Master Thesis

Predictive Analytics of Defective Machinery Parts in Reverse Supply Chain: A Case Study at ASML

Bartosz Szarszewski

TU Delft

ASML

Master Thesis

Predictive Analytics of Defective Machinery Parts in Reverse Supply Chain: A Case Study at ASML

Thesis report

by

Bartosz Szarszewski

to obtain the degree of Master of Science
at the Delft University of Technology
to be defended publicly on Monday 26th August, 2024 at 10:15

Thesis committee:

University chair:	Prof. dr. R.R. Negenborn | TU Delft ME
University supervisors:	Dr. J.M. Vleugel | TU Delft CiTG
	Dr. ir. M.B. Duinkerken | TU Delft ME
External supervisor:	Judith van Peursen MSc | ASML RSC Ops
Project duration:	March, 2024 - August, 2024
Master program:	Transport, Infrastructure & Logistics (TIL)
Date:	Tuesday 20th August, 2024

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

Faculty of Civil Engineering & Geosciences · Delft University of Technology

Preface

This thesis examines the challenges of managing the return flow of defective machinery parts in the semiconductor industry. As part of my master's program to complete my degree in Transport, Infrastructure, and Logistics at Delft University of Technology, I conducted this research using ASML's data, focusing on their return flows to develop predictive models aimed at improving the efficiency of their reverse supply chain operations while addressing key gaps in existing research. The study adapts and optimizes forecasting models specifically for reverse supply chains, determining which models are best suited for this sector.

I want to express my sincere thanks to ASML Netherlands and the Reverse Supply Chain Operations department for giving me the opportunity to learn from their operations and understand how such a complex supply chain works in reality. I am especially grateful to Judith van Peursen for her excellent supervision during my six-month thesis period. Her support made it much easier for me to understand the business. I also want to thank Sophie Stalpers, the manager of the team, for her helpful guidance and the opportunity to present my work to the teams on several occasions.

I would also like to extend my deep appreciation to my TU Delft graduation committee. Dr. Jaap Vleugel, my daily supervisor, has been a consistent source of support and guidance throughout this process, helping me approach the research from different perspectives. I am also grateful to Dr.ir. Mark Duinkerken for his valuable input in helping me structure my thesis efficiently and making critical statistical decisions. Lastly, I want to thank Prof.dr. Rudy Negenborn, the chair of the committee, for his role in organizing our meetings and for the constructive feedback he provided during our sessions.

Finally, I want to thank my family for their constant support throughout my studies. Although they did not have the knowledge to provide the specific insights I needed for my thesis, they always encouraged me to work hard and helped me to achieve my goals. Their support has been essential during my bachelor's years in Groningen and my pre-master and master's studies at TU Delft. I am especially thankful to my sister and girlfriend for their support during the challenging times of my thesis, and to Yannick for sharing his knowledge in machine learning and data analysis with me.

Bartosz Szarszewski
Amsterdam, August 2024

Abstract

Accurate forecasting in Reverse Supply Chain (RSC) management is crucial for the semiconductor industry, particularly for companies like ASML, which must efficiently manage the return flow of defective machinery parts. This study addresses key gaps by developing and evaluating time series-based forecasting models tailored to ASML's RSC. Using a modified ABC-analysis, parts were categorized based on defect frequency and economic impact, focusing on the most critical components. The research applied and optimized models including SES, ARIMA, ARIMAX, and LSTM, using five years of historical defect data. The analysis showed that LSTM models excel in high-frequency (weekly) forecasts for parts with frequent early-life defects, achieving an average mMAPE of 26.32%. ARIMAX models performed best for lower-frequency (monthly) data, particularly in sparsely represented classifications, with mMAPE as low as 4.31% to 10.80%. Despite a higher mMAPE of 85.42% in one outlier, ARIMAX emerged as the most balanced model, offering a practical trade-off between accuracy and computational efficiency. Furthermore, the study highlights the computational efficiency of ARIMAX, which, although more demanding than SES, provided a favorable balance, with ARIMA and LSTM being more resource-intensive. These findings demonstrate ARIMAX's suitability for long-term forecasting and broader trend analysis, making it the preferred model for ASML's RSC. This research provides a robust, data-driven framework that enhances inventory management and capacity planning, making it possible to predict return flows with low error rates on a monthly basis, particularly in high-tech industries where many defective parts are returned. Future research should incorporate additional dynamic variables, explore hybrid models, and refine data splitting techniques to further improve predictive accuracy and support sustainable supply chain operations.

Acronyms

11NC	11-digit Numerical Code
12NC	12-digit Numerical Code
ACF	Auto-Correlation Function
ADI	Average Demand Interval
ADF	Augmented Dickey-Fuller
AIC	Akaike Information Criterion
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Networks
AR	Auto-Regressive
ARIMA	AutoRegressive Integrated Moving Average
B2B	Business-to-Business
B2C	Business-to-Customer
CLSC	Closed-Loop Supply Chain
CSV	Comma-Separated Value
CV^2	Squared Coefficient of Variation
DLM	Distributed Lag Models
DMADE	Define, Measure, Analyze, Design, and Evaluate
DUV	Deep Ultraviolet
ERD	Entity Relationship Diagram
EUV	Extreme Ultraviolet
EWMA	Exponentially Weighted Moving Average
I	Integration component
LASSO	Least Absolute Shrinkage and Selection Operator
LSTM	Long Short-Term Memory
MA	Moving Average
MN	Material Notification
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
mMAPE	Modified Mean Absolute Percentage Error
ML	Machine Learning
MRB	Material Review Board
MRT	Material Reservation Tool
MSE	Mean Squared Error

MQ	Material Quality
OEM	Original Equipment Manufacturer
PACF	Partial Auto-Correlation Function
P&D	Planning and Delivery
PRM	Product Recovery Management
RCA	Root Cause Analysis
RNN	Recurrent Neural Networks
RL	Reverse Logistics
RMSE	Root Mean Squared Error
RSC	Reverse Supply Chain
RSC Ops	Reverse Supply Chain Operations
SAP	System Applications and Products in Data Processing
SARIMA	Seasonal ARIMA
SCM	Supply Chain Management
SES	Simple Exponential Smoothing
SMA	Simple Moving Average
SMAPE	Symmetric Mean Absolute Percentage Error
SVM	Support Vector Machine
SVR	Support Vector Regression
WEEE	Waste of Electrical and Electronic Equipment

Contents

Preface ii

Abstract iii

Acronyms iv

List of Figures viii

List of Tables x

1 Introduction 1
- 1.1 Company Analysis . . . 1
- 1.2 Research Context . . . 2
- 1.3 Problem Statement . . . 3
- 1.4 Research Objective . . . 3
- 1.5 Research Questions . . . 4
- 1.6 Scope . . . 4
- 1.7 Research Approach . . . 5
- 1.8 Thesis Outline . . . 7

2 Literature Research: Reverse Supply Chain Challenges 8
- 2.1 Reverse Supply Chain Dynamics . . . 8
- 2.2 Challenges and Importance of Forecasting in RSC . . . 11
- 2.3 Failure Rate Stages & Reliability . . . 12
- 2.4 Inventory Control & Classification . . . 14
- 2.5 Conclusion & Discussion . . . 17

3 Literature Research: Forecasting Methods & Metrics 18
- 3.1 Forecasting Methods & Models . . . 19
- 3.2 Evaluation Metrics . . . 27
- 3.3 Conclusion & Discussion . . . 30

4 System Analysis 32
- 4.1 CLSC & RSC Process Identification . . . 32
- 4.2 Part Identification and Notification Systems . . . 33
- 4.3 Swimlane Process Analysis . . . 36
- 4.4 Current vs Forecast-Driven Approach Comparison . . . 38
- 4.5 Conclusion . . . 38

5 Data Analysis 40
- 5.1 Data Systems & Focus . . . 40
- 5.2 Data Gathering . . . 41
- 5.3 Data Preparation . . . 43
- 5.4 Data Interpretation . . . 45
- 5.5 Parts Classification . . . 48
- 5.6 Classified Parts Analysis . . . 52
- 5.7 Data Transformation . . . 53
- 5.8 Conclusion . . . 56

6 Forecasting Models for Defective Machinery Parts 58
- 6.1 Requirements . . . 58
- 6.2 Variables Selection & Feature Engineering . . . 59
- 6.3 Selection of Forecasting Models . . . 59
- 6.4 Evaluation Metrics . . . 61
- 6.5 Conclusion . . . 63

7 Experiments & Results **64**
7.1 Simple Exponential Smoothing . . . 64
7.2 ARIMA Model . . . 66
7.3 ARIMAX Model . . . 70
7.4 LSTM Model . . . 72
7.5 Computational Effort of Models . . . 75
7.6 Conclusion . . . 77

8 Models Evaluation **78**
8.1 Models Comparison . . . 78
8.2 Practical Model Evaluation . . . 81
8.3 Conclusion . . . 82

9 Conclusion & Discussion **83**
9.1 Conclusion . . . 83
9.2 Discussion . . . 84

10 Recommendations **87**

References **92**

A Scientific Paper **93**

B Swimlane **102**

C Algorithms **104**
C.1 Grid Search for SES Parameters . . . 104
C.2 Grid Search for ARIMA Parameters . . . 105
C.3 Grid Search for ARIMAX Parameters . . . 106
C.4 Random & Grid Search for LSTM Parameters . . . 107

D Graphs **108**
D.1 Resampled Distribution Graphs . . . 108
D.2 Weekly & Monthly Forecasting Results per Model . . . 110
D.3 ACF, PACF, and ADF for each Part_Freq . . . 114
D.4 Weekly Forecasting Results Models Comparison . . . 126
D.5 Monthly Forecasting Results Models Comparison . . . 127

List of Figures

1.1 ASML's Systems Overview (ASML Holding N.V., 2015) 1
1.2 Forward and Reverse Supply Chain of ASML; Scope of This Research 5
1.3 CRISP-DM Model by Chapman et al. (2000) 6
1.4 Thesis Framework including Chapters, Research Questions, and Methodologies 7

2.1 Key processes of Reverse Logistics (Agrawal et al., 2015) 9
2.2 CLSC model integrating forward and reserve SC (Guide and Van Wassenhove, 2001) . . . 9
2.3 Integrated Supply Chain Model (Thierry et al., 1995) 10
2.4 Bathtub Curve (Lewis, 1996) 13
2.5 Diagram of Logistics Control Processes and Decisions (Driessen et al., 2014) 15
2.6 Demand Pattern Classification (Syntetos and Boylan, 2005) 16

3.1 Overview of Forecasting Methods & Models Discussed (Authors' own creation) 18
3.2 An Adaptive Neuro-Fuzzy Inference System (ANFIS) architecture by Jin et al. (Aug. 2013) . 26
3.3 An Long Short-Term Memory (LSTM) cell (Ismail et al., 2018) 27
3.4 Matrix of Forecasting Models in Literature Research (Authors' own creation) 31

4.1 General CLSC of ASML (Authors' own creation) 32
4.2 Forward Flow of ASML Supply Chain (Authors' own creation) 33
4.3 Reverse Flow of ASML Supply Chain (Authors' own creation) 34
4.4 12NC, Equipment Number and Material Notification (MN) (Authors own creation) 34
4.5 Material Notification Creation (ASML's Process Modified by Author) 36
4.6 Swimlane Process of Defective Parts (Authors' own creation) 37

5.1 Data Analysis Process Overview (Authors' own creation) 40
5.2 Connection Between Data Systems (Authors' own creation) 41
5.3 Entity Relationship Diagram of Data Merging (Authors' own creation) 43
5.4 Total defects aggregated per month over the last 5 years (Author's Own Creation) 46
5.5 Early Life Defects Pareto (Author's Own Creation) 47
5.6 Top 5 Detection Code Descriptions (Author's Own Creation) 48
5.7 Top 5 Source of Cause Code Descriptions (Author's Own Creation) 48
5.8 Defective Parts Notifications Pattern (Authors' own creation) 49
5.9 ABCD-analysis Histogram with Pareto (Authors' own creation) 50
5.10 Unique Parts Defect Distribution (Authors' own creation) 53
5.11 Unique Parts Cumulative Distribution (Authors' own creation) 54
5.12 Part A2 Resampled Distribution (Left: Weekly; Right: Monthly) 54

6.1 Selected Models Matrix from Literature Research (Authors' own creation) 60

7.1 Data Extraction & Transformation connection with Experiments & Results 64
7.2 Weekly and Monthly SES Results for Part C (rank 1 & 2) and Part A1 (rank 9 & 12) 66
7.3 ACF and PACF plots for original defect counts of Part A1_Wk 67
7.4 ACF and PACF plots for first differenced defect counts (d=1) of Part A1_Wk 68
7.5 Weekly and Monthly ARIMA Results for Part C (rank 1 & 2) and Part A1 (rank 9 & 12) . . . 69
7.6 Weekly and Monthly ARIMAX Results for Part C (rank 2 & 3) and Part A1 (rank 8 & 12) . . 71
7.7 Weekly and Monthly LSTM Results for Part C (rank 1 & 10) and Part A1 (rank 8 & 12) . . . 76

B.1 Swimlane Process of Defective Parts (Authors' own creation) 103

D.1 Weekly Resampled Parts Distribution 108
D.2 Monthly Resampled Parts Distribution 109

D.3 Weekly and Monthly SES Results . . . 110
D.4 Weekly and Monthly ARIMA Results . . . 111
D.5 Weekly and Monthly ARIMAX Results . . . 112
D.6 Weekly and Monthly LSTM Results . . . 113
D.7 ACF, PACF, and ADF for Part A1_Wk . . . 114
D.8 ACF, PACF, and ADF for Part A1_Mh . . . 115
D.9 ACF, PACF, and ADF for Part A2_Wk . . . 116
D.10 ACF, PACF, and ADF for Part A2_Mh . . . 117
D.11 ACF, PACF, and ADF for Part A3_Wk . . . 118
D.12 ACF, PACF, and ADF for Part A3_Mh . . . 119
D.13 ACF, PACF, and ADF for Part B1_Wk . . . 120
D.14 ACF, PACF, and ADF for Part B1_Mh . . . 121
D.15 ACF, PACF, and ADF for Part B2_Wk . . . 122
D.16 ACF, PACF, and ADF for Part B2_Mh . . . 123
D.17 ACF, PACF, and ADF for Part C_Wk . . . 124
D.18 ACF, PACF, and ADF for Part C_Mh . . . 125
D.19 Weekly SES, ARIMA, ARIMAX, LSTM Results . . . 126
D.20 Monthly SES, ARIMA, ARIMAX, LSTM Results . . . 127

List of Tables

3.1 Comparison of Simple Forecasting Methods 21
3.2 Comparison of Exponential Smoothing Methods 22
3.3 Comparison of Autoregressive and Related Forecasting Methods 24
3.4 Comparison of Supervised Learning Methods 25
3.5 Comparison of Deep Learning Methods 28
3.6 Evaluation of forecasting models across different metrics, categorized by cited studies. . . 30

4.1 Comparative Analysis of Current vs. Forecast-Driven Approaches at ASML 39

5.1 Summary of Data Columns with Non-Null Counts, Null Counts, and Unique Counts 44
5.2 Summary of Missing Values Steps 45
5.3 Summary of Cleaned Data Columns with Non-Null Counts, Null Counts, and Unique Counts 46
5.4 Initial Classification Results 49
5.5 Ranges for Each Category (Normalized) 51
5.6 ABCD-Categorically Aggregated Data 51
5.7 Top 6 Highest 12NC Parts per Category (Normalized) 52
5.8 Statistics for Each Unique Part 52
5.9 Weekly Statistics for 80/20 Split 56
5.10 Monthly Statistics for 80/20 Split 56
5.11 Part Abbreviations 56

6.1 Selected Evaluation Metrics 62

7.1 Optimal Parameters and Performance Metrics for SES 65
7.2 Parameter Ranges for ARIMA Model Grid Search 68
7.3 Optimal Parameters and Performance Metrics for ARIMA Model 69
7.4 Optimal Parameters and Performance Metrics for ARIMAX Model 71
7.5 Pearson Correlation Coefficients for Exogenous Variables (ARIMAX) 72
7.6 Parameter Ranges for Random Search 73
7.7 Selected Hyperparameter Ranges 74
7.8 Selected Window Size Ranges 74
7.9 Optimal Parameters and Performance Metrics for LSTM Model 75
7.10 Pearson Correlation Coefficients for Exogenous Variables (LSTM) 76
7.11 Computational Time of Grid Searches *(in seconds)* 76

8.1 All Models and Part Frequencies Performance Metrics 79
8.2 Average A & B Part Classification All Models and Part Frequencies Performance Metrics . 80
8.3 Overall Average Performance Metrics for Each Model (Wk and Mh) 81

1
Introduction

This chapter introduces the study by providing an overview of an company analysis of ASML, highlighting the research context, and formulating the problem statement. It outlines the research objectives and questions, details the methodologies to be used, and concludes with a brief overview of the thesis structure.

1.1. Company Analysis

ASML, a Dutch multinational, plays a key role in the semiconductor industry as an Original Equipment Manufacturer (OEM). It specializes in developing and manufacturing sophisticated photolithography machines that are integral for producing nanometer-thin silicon wafers. These wafers are crucial for the microchip production utilized by major technology firms such as Samsung, TSMC, and Intel for manufacturing a range of electronic devices, from laptops to mobile phones. Consequently, ASML is positioned at the heart of a vital supply chain, essential for the advancement of technology in global tech enterprises (The Economist, 2024).

Headquartered in Veldhoven, Netherlands, ASML's history is rooted in its establishment in 1984, resulting from a collaboration between Philips and ASMI in Eindhoven. Over the years, ASML has evolved into a significant force within the semiconductor sector, marked by its expansion across over 60 locations in 16 countries. Its technological milestones include the introduction of the PAS 5500 series, the development of immersion technology in the TWINSCAN systems, and the pioneering efforts in Extreme Ultraviolet (EUV) lithography. With a keen focus on EUV and Deep Ultraviolet (DUV) technologies, ASML not only dominates the DUV market but also enjoys a unique monopoly in EUV lithography, offering systems that cater to a broad spectrum of semiconductor manufacturing requirements. The chronological development of the systems, including their specifications, are illustrated in Figure 1.1.

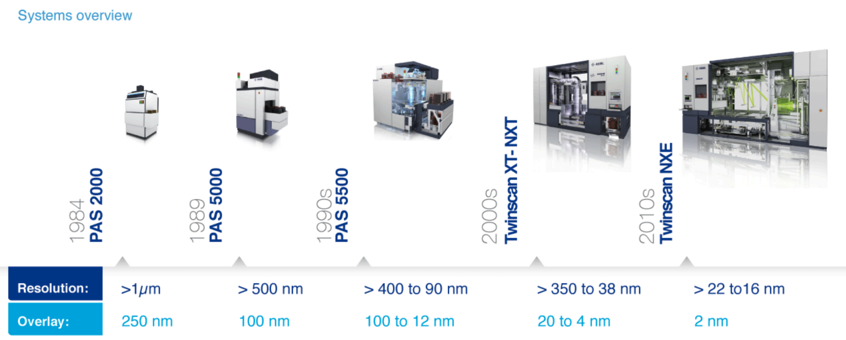


Figure 1.1: ASML's Systems Overview (ASML Holding N.V., 2015)

ASML's commitment to environmental sustainability is demonstrated through its goal to achieve net-zero greenhouse gas emissions and its efforts to reduce waste to landfill or incineration by 2030. The company's ambitious target of a 95% reuse rate for defective machine parts, which are returned from the field or from their own factories, by 2025 illustrates its dedication to integrating environmental responsibility with economic benefits through sustainable practices (ASML, 2022).

The company's operational success is further bolstered by its comprehensive Supply Chain Management (SCM) strategy. With over 400 suppliers predominantly based in Europe, ASML ensures a seamless flow of machinery parts from these suppliers to its manufacturing facilities and then to the customers. This end-to-end support system encompasses the provision of service parts for global customers, facilitating the swift replacement of defective parts.

This thesis encompasses collaboration with ASML's Reverse Supply Chain Operations (RSC Ops) team, a crucial segment of the Planning and Delivery (P&D) department. The P&D department plays a vital role in aligning ASML's strategic goals with its capacity planning initiatives. Particularly, the RSC Ops team is dedicated to the analysis, implementation, and improvement of Reverse Supply Chain (RSC) processes. These processes manage the logistics of returning products from customers to the warehouse, factory, and both external and internal suppliers, including returns from ASML's main factory in Veldhoven to its storage facilities and suppliers. The primary objective of these operations is to quickly assess and execute the necessary actions for these returned parts, which include repair, component harvesting, dismantling, and recycling of material components.

1.2. Research Context

The RSC is integral to achieving sustainability and operational efficiency within modern supply chains, particularly in resource-intensive industries like the semiconductor sector. Current practices emphasize the strategic integration of RSC into traditional SCM frameworks, showcasing the importance of Closed-Loop Supply Chain (CLSC) strategies. These strategies integrate forward and reverse supply chains, enhancing environmental objectives through extended product lifecycles and minimized waste, particularly in sectors such as automotive and electronics where sector-specific RSC strategies are crucial (Guide and Van Wassenhove, 2001; Kumar and Putnam, 2008; Bressanelli et al., 2019).

Despite these advances, several challenges persist, notably in forecasting and data management within the RSC. Literature indicates a significant scarcity of data on return volumes, hindering effective management (Toktay et al., 2003; Cui et al., 2020). Moreover, there is a pronounced lack of methodological comparisons and empirical validations of forecasting models tailored for the RSC (Kumar et al., 2014; Syntetos et al., 2016). Such gaps hinder the development of accurate forecasting models crucial for planning and optimizing return flows. The primary focus remains on demand forecasting, with insufficient attention to the supply-side forecasting of returned defective parts. This oversight undermines the accuracy of predicting and managing returns.

Furthermore, the strategic relevance of Product Recovery Management (PRM) within the manufacturing industry is underscored in the integrated supply chain model of Thierry et al. (1995), which highlights the importance of effectively integrating recovery operations such as repair, refurbishing, and remanufacturing back into the forward supply chain. This facilitates the efficient reuse and recycling of high-value parts, reducing waste and enhancing resource efficiency.

In terms of reliability and quality control, the intersection of failure rates with the return of defective parts in semiconductor manufacturing is critically examined. Traditional models like the bathtub curve and Weibull distribution are effective in later stages but fall short during initial phases where defects are prevalent (Abernethy, 2006; Roesch, 2012). This suggests a need for alternative statistical methods such as AutoRegressive Integrated Moving Average (ARIMA) and machine learning techniques for more accurate early-stage predictions (Lee et al., 2021).

This research seeks to address these identified gaps by exploring the specific challenges and opportunities of RSC operations in the semiconductor industry. By examining the strategic integration of RSC within SCM frameworks, assessing the effectiveness of current forecasting models, and exploring innovative predictive techniques, the study aims to enhance the understanding and management of RSC operations. It specifically focuses on tailoring forecasting approaches to the unique needs of the semiconductor sector for defective machinery parts early in the product lifecycle, thereby aligning with corporate sustainability goals and operational demands. This approach not only contributes to the broader discourse on efficient and sustainable supply chain practices but also fosters the development of sector-specific solutions.

1.3. Problem Statement

ASML is confronting substantial challenges within its RSC operations, particularly concerning the inventory of returned defective parts needing repair or upgrading. The current state of the reverse inventory is alarmingly overstocked, holding more than 100,000 returned parts. This inventory issue stems primarily from the complex process flows within the RSC, where parts undergo evaluations to determine whether they should be dismantled, harvested, repaired, or recycled.

A significant challenge arises when these parts accumulate in storage for extended periods due to an initial lack of demand, compounded by previous investments in newly purchased parts. This backlog not only ties up substantial financial resources but also disrupts operational fluidity, hindering sustainability and efficiency objectives within the RSC.

This problem is particularly severe in the reverse flow of parts from ASML's factory located in Veldhoven, where parts are assembled into new modules but may fail during the assembly process, particularly in the infant stage of the product lifecycle. These parts must await a demand signal before they can be returned to the supplier for repair and subsequently reused as spare parts. Due to capacity constraints and prolonged storage, some parts are eventually harvested or recycled, which substantially diminishes their value and circularity.

Moreover, a recent policy change mandates that all repairable defective parts originating from customer facilities be returned for repair to the specific supplier, regardless of immediate demand. These parts are repaired and stored in a different inventory than the reverse inventory, and therefore, the scope of defective parts from customers is not the main problem of the RSC Ops.

Currently, ASML's RSC lacks a forecasting model to provide insights into future inflows of defective parts into the reverse inventory. This gap results in excess and obsolescence of parts, leading to significant value loss. Often, repairable defective parts must wait due to earlier orders for new parts. Implementing a forecasting model to predict the inflow of repairable defective parts could enable the early blocking of new purchases, allowing these parts to be stored for shorter periods or sent directly to the supplier for repair from the Veldhoven factory.

A forecasting model could also be utilized for short-term predictions on a monthly basis, aligning with ASML's practice of updating KPI's and inventory numbers at the beginning of each month. This alignment ensures that forecasts remain relevant and integrated with overall operational planning. Furthermore, short-term predictions are valuable for the RSC Ops team, which reviews part flows to maintain smooth operations.

There is a need for a predictive forecasting model tailored to the unique dynamics of the semiconductor sector's RSC. Developing a forecasting model is essential for optimizing the use of repaired parts, reducing dependency on new part production, and alleviating financial strain caused by overstocked inventories. This development could enhance the operational management of ASML's RSC, aligning it more closely with the company's strategic objectives and contributing to the sustainability and efficiency of the semiconductor industry's SCM.

1.4. Research Objective

The primary research objective of this study is outlined below.

Research Objective

To examine forecasting models that accurately predict the return flow volume of defective machinery parts.

This objective focuses on utilizing ASML's operational data to investigate and modify predictive forecasting models tailored to address ASML's specific challenges. Furthermore, these models aim to be generalizable across the semiconductor industry to enhance RSC operations.

The objective will be explored through the following sub-objectives:

- **Context Definition**: Examine existing methods and challenges in forecasting the return of defective machinery parts within the RSC. This includes identifying key variables, suitable forecasting methods, and evaluation metrics to document challenges and explore improvements applicable to the semiconductor industry.

- **Current State Assessment**: Evaluate ASML's current RSC management practices by reviewing CLSC dynamics and focusing on reverse inventory management and supply chain efficiency. Identify inefficiencies and areas for potential improvement.
- **Data and Variable Analysis**: Analyze ASML's historical data to pinpoint critical variables influencing the return quantities of defective machinery parts. Conduct a detailed examination of historical trends to inform the selection of an effective forecasting model.
- **Models Experimentation**: Select and test forecasting models to predict the future return quantities of defective machinery parts at ASML. The goal is to identify a model that balances computational efficiency with accuracy, enhancing the management of returns. This model should potentially be used for blocking new purchases or for monthly operational planning purposes.
- **Models Evaluation**: Apply thorough testing and assessment to processed ASML data and specialised software. Evaluate the forecasting model's accuracy, operational efficacy, scalability, and broader applicability in the semiconductor sector to improve overall operational efficiency.

1.5. Research Questions

This section introduces the research questions designed to assess the research objectives. The main question addressed below.

Main Research Question

How can forecasting models accurately predict the return flow volume of defective machinery parts?

The following sub-questions delve into various aspects of this main question:

1. What are the existing practices and challenges in managing the reverse supply chain for defective machinery parts?
2. Which forecasting methods and models are available and suitable for predicting defective machinery parts in the return flow, and what are the crucial variables and evaluation metrics?
3. What are the key challenges in managing the return flow of defective parts within ASML's reverse supply chain?
4. What data is available at ASML for predicting the return flow volume of defective parts, and how can this data be processed with a focus on the crucial variables?
5. Which forecasting models and evaluation metrics are best suited for predicting the return flow of defective machinery parts at ASML, considering the specific requirements and crucial data variables?
6. What are the optimal parameters for the most suitable forecasting models tailored to different classifications of defective parts and varying data frequencies?
7. Which forecasting models provide the highest accuracy in predicting defect counts for specific classifications of defective machinery parts, considering various data frequencies?

1.6. Scope

The scope of this master's thesis is clearly defined to ensure the research remains focused and manageable. Precision in defining the scope is crucial to the study's success since it streamlines the research process and ensures that the questions addressed are adequately answered. The following important aspects are considered when determining the scope of this thesis.

- **Focus on the RSC Framework:** The thesis specifically addresses the return flow of defective machinery parts within the RSC framework. Following Driessen et al. (2014), two key decision points are highlighted: the strategic or tactical classification of parts concerning their returns, usually conducted annually, and more frequent tactical forecasts (monthly or quarterly) to predict return rates and times. Although inventory control and demand forecasting are not the main objectives of this research, their connection with the return forecasting model is crucial. An effective and integrated forecasting model must consider these aspects to ensure alignment with the broader supply chain operations. Therefore, while inventory control and demand forecasting are not directly studied, their influence on the returns forecasting model is acknowledged to ensure the developed model functions effectively within ASML's operational framework.

- **Industry-Specific Focus:** This research is rooted in the OEM sector, focusing predominantly on Business-to-Business (B2B) interactions while placing less emphasis on Business-to-Customer (B2C) dynamics. As described in the research context the case study will be focused on the semiconductors industry. To strengthen the study, an extensive literature review spanning various industries is conducted, incorporating a diverse range of forecasting methods, variables, and evaluation metrics.
- **Geographic and Process-Level Focus at ASML:** The operational focus of this research is detailed in Figure 1.2. This diagram illustrates the flow of parts from OEM suppliers through the ASML factory in Veldhoven (the Netherlands), where parts are assembled into modules. The primary focus of this research is shown within the blue rectangle in the diagram. This area highlights the initial reverse flow of parts, from the detection of defects at the manufacturing site through to the decision-making regarding repairs, depicted by the orange lines. Concentrating on this specific scope is essential as it allows for a precise analysis of defect management processes. By focusing on the initial stages of the reverse flow, the research aims to develop an accurate forecasting model for defective parts at ASML. This targeted approach ensures that the study addresses the most critical phase of defect handling, providing valuable insights for improving inventory management and operational efficiency.
- **Product-level focus at ASML:** ASML's operations encompass a diverse array of machinery parts, each prone to defects. To systematically identify the specific parts for this study, a classification analysis method will be employed, as highlighted as a crucial strategy in the literature review. Such an analysis should categorize parts based on the variability in defect occurrences and their economic value, ensuring a balanced and representative selection. This targeted selection ensures that the study focuses on parts that significantly impact ASML's operations, excluding end-of-life parts and concentrating on those in the early stages of their lifecycle. By prioritizing parts with the highest economic impact and varied defect ranges, the research aims to provide robust insights into defect forecasting. The economic calculation of the parts' impact on inventory is not part of this research; instead, the focus is on determining the volume of defective parts using the most accurate forecasting model.

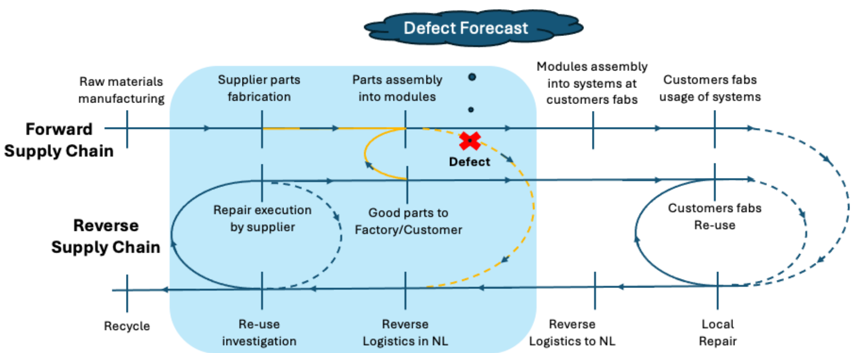


Figure 1.2: Forward and Reverse Supply Chain of ASML; Scope of This Research

1.7. Research Approach

For a master's thesis, the research methodology must align closely with the study's objectives and complexities. This study aims to examine forecasting models that can accurately predict the return flow volume of defective parts for managing the RSC at ASML. Initially, the Define, Measure, Analyze, Design, and Evaluate (DMADE) methodology, a modification of DMADV with the Verify phase replaced by Evaluate, was proposed to better fit the iterative nature of academic research. However, as the research progressed, a stronger emphasis on data management led to the adoption of the CRISP-DM methodology outlined by Chapman et al. (2000). CRISP-DM, the Cross-Industry Standard Process for Data Mining, starts with the Business Understanding phase. This step ensures that the research objectives align with business needs and identifies the critical processes involved. The subsequent phases include Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. This systematic approach ensures coherent development and practical implementation of the forecasting model at ASML. The stepwise phases of the methodology are displayed in Figure 1.3.

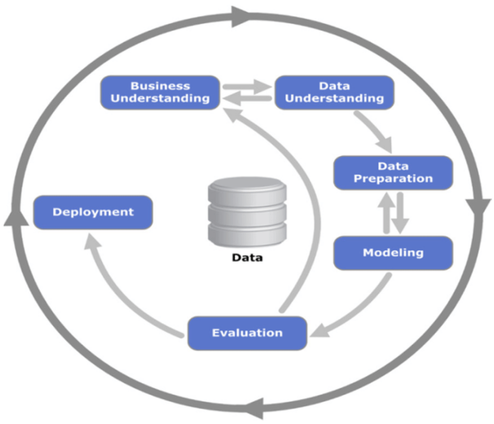


Figure 1.3: CRISP-DM Model by Chapman et al. (2000)

The research starts with a literature review followed by a system analysis, similar to the Business Understanding phase in CRISP-DM, to ground the study in theory and practice. Data Understanding and Preparation phases are critical, ensuring data integrity for modeling. The Modeling phase applies statistical and data mining methods, followed by a precise evaluation by comparing the models to each other. This process aims to identify the models that offer the highest prediction accuracy. The Deployment phase, which focuses on implementing and monitoring the model at ASML, is not part of this thesis.

By integrating CRISP-DM, the research bridges theoretical research and practical application, ensuring outcomes are both academically sound and practically actionable for ASML's RSC management.

The research methodologies include:

- **Literature Research**: This phase follows a structured review process, beginning with the formulation of the research problem. An extensive search of relevant literature is conducted, with careful screening for relevance and quality. The literature is then extracted, analyzed, and synthesized to form a robust theoretical foundation. This foundational knowledge informs the first two sub-questions, focusing on the core challenges, best practices in RSC management and forecasting models.
- **Desk Research & Expert Consulting**: Complementary to the literature review, this phase involves gathering additional data from various sources such as industry reports and process documentation. Expert consultations are conducted to validate findings and provide practical insights, ensuring the research remains grounded in real-world practices and challenges.
- **Swimlane Diagram**: The Swimlane diagram methodology is employed to systematically map the RSC processes at ASML. This visual tool helps in understanding the operational framework by clearly delineating the roles, responsibilities, and interactions between different departments. The Swimlane diagram serves as a foundational map that guides the data understanding and preparation by highlighting key process flows and outputs.
- **Data Preparation**: This crucial phase involves the collection and initial processing of raw data from internal sources at ASML. The data is intensively and precisely gathered to ensure it accurately represents the system's functionality and challenges. Following collection, the data undergoes a preparation process to ensure quality and consistency, making it suitable for in-depth analysis.
- **Data Understanding**: In this phase, the prepared data is thoroughly analyzed to gain insights into the operational patterns and relationships within the RSC. This step is essential for identifying crucial variables that will be used as inputs for the forecasting models. Understanding the data's structure and content is critical for the subsequent modeling phase.

- **Requirements & Selection Process**: Based on the insights gained from the data understanding phase, specific requirements for the forecasting models are defined. This involves selecting appropriate models and techniques that best fit the characteristics of the data and the operational needs of ASML's RSC. The selection process is guided by both theoretical considerations and practical constraints.
- **Parameters Optimization**: Once the models are selected, they undergo a complex optimization process. This involves adjusting various parameters to enhance the models' predictive accuracy and reliability. The optimization process ensures that the models are finely tuned to the specific context of ASML's RSC.
- **Models Evaluation**: The final phase involves a comprehensive evaluation of the forecasting models. The models are tested against historical data to assess their predictive performance. Various evaluation metrics will be used to measure accuracy, reliability, and relevance to ASML's strategic objectives. This phase ensures that the models not only perform well theoretically but also provide actionable insights for practical implementation.

1.8. Thesis Outline

This thesis is carefully structured to ensure efficient research progression. The conceptual framework of the thesis is illustrated in Figure 1.4, which delineates the methodologies employed in each chapter and maps each chapter to its corresponding research question.

The process initiates with a literature research in Chapters 2 and 3, addressing the foundational aspects of the study. Subsequent chapters build upon this groundwork with desk research, expert consulting, data preparation, and understanding, leading to the selection, optimization and evaluation of forecasting models. Each chapter systematically addresses specific sub-questions, culminating in Chapter 9, where the main research question is thoroughly examined, followed by recommendations in Chapter 10.

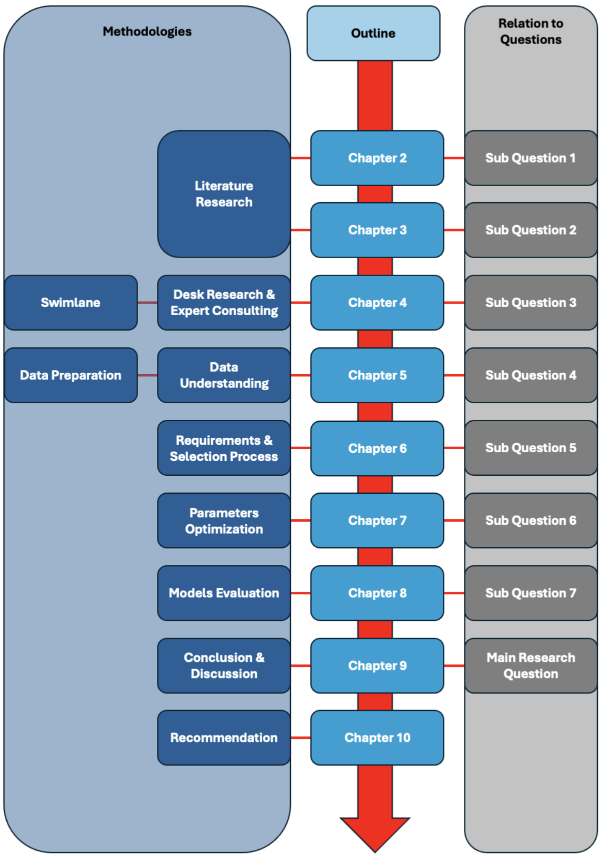


Figure 1.4: Thesis Framework including Chapters, Research Questions, and Methodologies

2

Literature Research: Reverse Supply Chain Challenges

This chapter presents a comprehensive literature review on the RSC within SCM, specifically focusing on the challenges related to forecasting the return flows of defective parts. It examines existing practices and methodologies, identifying both opportunities and difficulties in managing the inventory and Reverse Logistics (RL) aspects of the RSC for defective machinery parts. By the end of this chapter, the first sub-research question of this thesis will be answered and discussed.

1. *What are the existing practices and challenges in managing the reverse supply chain for defective machinery parts?*

In conducting this literature review, a structured search strategy was employed to ensure a comprehensive examination of relevant academic and industry sources. The initial search terms used included ”reverse supply chain”, ”reverse logistics”, ”defective parts”, ”remanufacturing”, ”circular economy”, ”inventory management”, and ”return flow”. These terms were chosen to cover a broad spectrum of RSC-related challenges and practices.

To further refine the search and include seminal works and recent advancements, the snowballing technique was utilized. This involved reviewing the reference lists of key articles identified in the initial search to discover additional relevant studies. This iterative process captured a wide range of perspectives and methodologies, enriching the literature base for this research.

2.1. Reverse Supply Chain Dynamics

SCM is an overarching framework that integrates all activities related to the supply chain, procurement, and logistics. At its core, SCM aims to manage the flow of goods and services efficiently from inception to delivery. Fahimnia et al. (2013) describe the traditional supply chain, often referred to as the open-loop or forward supply chain, which focuses on the seamless progression of products and services from suppliers to manufacturers, then to retailers, and ultimately to end-users. This flow is facilitated by a network of transportation and storage operations, ensuring that goods move smoothly and efficiently through each phase of the supply chain.

Building on this foundational understanding, Fahimnia et al. (2013) extend the discussion to a mixed integer non-linear model for production-distribution planning, utilizing a customized algorithm and leveraging real-world data from the automotive industry. This model highlights the strategic integration of production and distribution processes, underscoring the importance of optimizing supply chain efficiency to accommodate both the traditional flow of goods and the complexities of modern supply chain demands.

The RSC is an critical component of modern SCM, focusing on the recovery and reutilization of products after consumer use. This emerging modern supply chain model diverges from conventional logistics by prioritizing the collection, RL, and reintroduction of products into the supply cycle (Guide and Van Wassenhove, 2006). The RSC serves not only to manage returns due to defects but also represents a strategic shift towards sustainable practices. OEM play a pivotal role in this process, balancing the production of new materials with the refurbishment and remanufacturing of existing products for further use (Guide and Van Wassenhove, 2001).

The return product flow in the RSC initiates with the collection of used or end-of-life products from consumers in a B2C supply chain. As described by Agrawal et al. (2015), this process involves several key steps, each presenting unique challenges and opportunities for value recovery. The journey of used or returned products is depicted in Figure 2.1, illustrating the pathway from collection, through inspection and sorting, to their eventual repair, remanufacturing, recycling, reuse, or disposal. This pathway is determined by whether the goal is to recapture value or to dispose of the product responsibly.

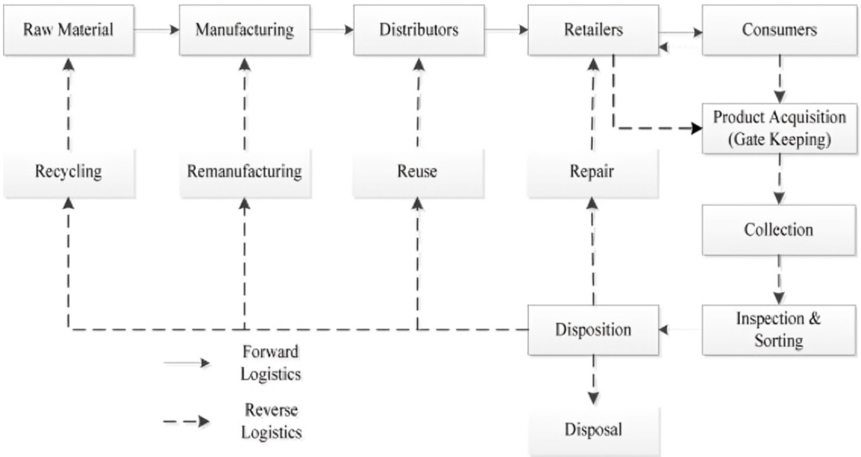


Figure 2.1: Key processes of Reverse Logistics (Agrawal et al., 2015)

The operational complexities of the RSC are highlighted by Diener (2004), who notes the issues of non-uniform product quality, the difficulty of forecasting return volumes, potential damage to product packaging, unclear destination or routing, inconsistent inventory management, and the less transparent nature of RL processes. Moreover, Fleischmann (2003), Ravi and Shankar (2014) emphasize the significant difficulties in anticipating return flows, exacerbated by the uncertainties inherent in the nature of returns. These uncertainties necessitate a robust and flexible RSC network to effectively manage the variability in return volumes, timing, and product condition.

The strategic importance of efficiently managing return product flows extends beyond logistical efficiency to encompass value creation through sustainability practices. Guide and Van Wassenhove (2006) delineates the RSC's role in collection, RL, inspection, sorting, and the recovery operations of used products, underscoring the value-driven focus of modern RSC management. This strategic perspective is further reinforced by the movement towards integrating the RSC within a circular economy framework, as discussed by Mishra et al. (2022). The circular economy model, supported by recycling, remanufacturing, reuse, and repair practices, aligns with the objectives of the RSC to extend product lifecycles and minimize waste. The significance of circular economy in the context of RSC is also supported by Bressanelli et al. (2019), advocating for the development of policies that encourage the transition to sustainable supply chain practices.

The concept of CLSC represents an advanced integration of forward and reverse supply chain activities, aimed at creating a seamless flow of materials and information. Guide and Van Wassenhove (2001) provides an illustration of a CLSC model in Figure 2.2, highlighting the integration of forward and reverse supply chains into a cohesive system. This integration facilitates a more sustainable and efficient management of resources, enabling the re-purposing of returned products in a manner that supports both economic and environmental objectives.

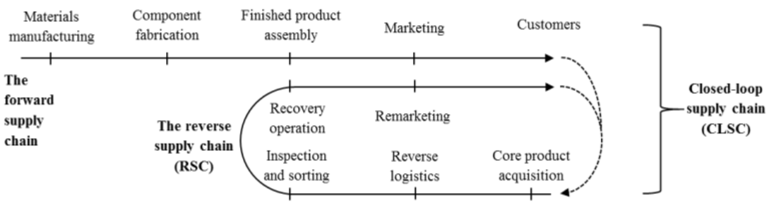


Figure 2.2: CLSC model integrating forward and reserve SC (Guide and Van Wassenhove, 2001)

The RSC strategies vary significantly across different sectors, necessitating tailored approaches and expert knowledge, especially given the potential outsourcing of RSC functions. In industries such as automotive and electronics, strategic RSC management is critical for achieving economic efficiency and addressing the unique challenges presented by the collection and reintegration of returned products (Kumar and Putnam, 2008).

Thierry et al. (1995) presents an integrated supply chain model in Figure 2.3, which effectively integrates PRM with traditional supply chain operations. This model is particularly relevant to the semiconductor industry, focusing on the post-fabrication stage where part outflows become inflows for subsequent module assembly, leading to the creation of complete semiconductor systems. The model delineates essential recovery processes such as repair, refurbishing, remanufacturing, cannibalization, and recycling, each necessitating varying degrees of disassembly and processing. By effectively integrating these recovery operations back into the forward supply chain, the model facilitates the efficient reuse and recycling of high-value components, thus reducing waste and enhancing resource efficiency within the industry's manufacturing processes (Thierry et al., 1995).

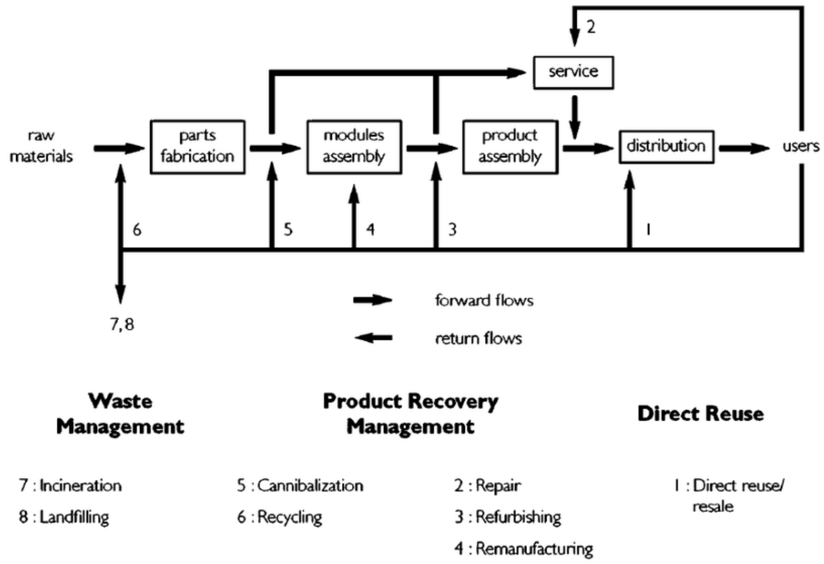


Figure 2.3: Integrated Supply Chain Model (Thierry et al., 1995)

Companies now view the RSC not just as a regulatory compliance mechanism but as a strategic tool capable of yielding economic advantages while bolstering corporate social responsibility. This strategic repositioning has spurred research and innovation, seeking to refine RSC operations and enhance corporate sustainability (Govindan et al., 2012).

2.1.1. Conclusion

This section delineates the strategic integration of RSC within traditional SCM frameworks, highlighting the necessity to manage return flows effectively for sustainability and operational efficiency. It emphasizes the role of CLSC in integrating forward and reverse logistics, thereby supporting environmental objectives and economic gains through extended product lifecycles and minimized waste (Guide and Van Wassenhove, 2001). Moreover, the literature discusses the importance of developing sector-specific RSC strategies to navigate the unique challenges and maximize opportunities within industries like automotive and electronics (Kumar and Putnam, 2008; Bressanelli et al., 2019). This strategic focus not only aligns with compliance but positions RSC as a crucial component for achieving corporate sustainability and leveraging economic advantages. The section advocates for ongoing research and tailored approaches in RSC operations to enhance efficiency and sustainability across different sectors (Guide and Van Wassenhove, 2006; Mishra et al., 2022).

2.2. Challenges and Importance of Forecasting in RSC

Forecasting the supply of returned parts within the RSC framework is important for enhancing the resilience and efficiency of supply chains. The literature reveals various forecasting models and methodologies, each contributing insights into addressing the inherent uncertainties of RSC processes. This part of the literature review synthesizes key challenges and used methodologies from the referenced studies, outlining their implications for future research in RSC forecasting.

Fleischmann (2003) and further discussions by Ravi and Shankar (2014) highlight the difficulties in forecasting return flows, which are intensified by the unpredictable nature of returns. They argue for the development of a strong RL network design, underlining the importance of creating models that are capable of adapting to variations in the amount and condition of returns. This involves a detailed look at the significance of inventory management and transportation logistics within the RL network, identifying it as a promising area for future academic research.

Furthermore, Thierry et al. (1995) highlight the crucial function of controlling entry and managing product recovery in guiding products' movement through RL channels. They argue that choosing the best ways to handle returned products, aimed at maximizing their recoverable value, is closely linked to the efficiency of forecasting models in predicting return volumes and timing. This connection underscores the necessity of developing forecasting models that not only predict the quantity of returns but also aid in strategic decisions about handling returns to optimize both economic benefits and environmental sustainability.

Agrawal et al. (2015) underscores the importance of precise forecasting in managing returns within RL processes. Given the inherent uncertainty in the volume, timing, and quality of returns, there is a pressing need for robust forecasting models to efficiently optimize RL operations. Despite this necessity, the literature, as highlighted by Toktay et al. (May 2003), reveals a notable lack of detailed forecasting practices in the RL as part of the RSC. This indicates a significant gap in both practical application and the development of methodologies specifically designed for RSC.

Toktay et al. (May 2003) examine the crucial role of accurate return flow forecasting in inventory control for remanufacturable products. They categorize products by lifecycle and upgrade potential, from consumables to durables, emphasizing the strategic, tactical, and operational decisions impacted by return flow predictions. The study introduces a methodological framework for forecasting returns, which involves developing a return delay model and refining its parameters based on historical sales and returns data. Their findings highlight the trade-offs between different forecasting methods in terms of cost performance and order variability, advocating for adaptable strategies tailored to various stages of product lifecycles and sales volumes.

Exploring the complexities of RSC, Kumar and Putnam (2008) identifies the variability and complexity of RL strategies across industries, suggesting that forecasting models must accommodate diverse sector-specific challenges. The anticipation of outsourcing RL activities further emphasizes the need for sophisticated forecasting models capable of navigating the unpredictability of return flows, highlighting the critical role of information systems and infrastructure in supporting effective RL processes.

Duc et al. (2010) and DeCroix et al. (2009) contribute insights into demand patterns and inventory optimization, relevant to both forward and reverse supply chains. Their findings suggest that integrating stochastic modeling and time-series analysis could offer valuable strategies for managing returns more effectively, pointing towards a need for forecasting models that better account for the stochastic nature of the supply in RSC.

Cui et al. (2020)'s research explores predictive models for estimating return volumes in a B2C supply chain, with a focus on the Least Absolute Shrinkage and Selection Operator (LASSO) method. Although LASSO is precise, Cui et al. (2020) points out its limitations in specifically identifying returns caused by product defects. This finding is critical as it highlights the need for more detailed models capable of understanding the varied reasons behind product returns. The study suggests that including factors like product type and manufacturing methods could lead to more accurate predictions in RSC.

Matsumoto and Komatsu (2015) and Thierry et al. (1995) contribute to the discussion on the effectiveness of forecasting models in RSC, particularly regarding the supply of return defective parts. Matsumoto and Komatsu (2015) provides an empirical investigation into the use of two different time series models for demand forecasting in auto parts remanufacturing, underscoring the significant potential of time series analysis in capturing seasonal and periodic variations in remanufactured product sales. This study's findings offer a benchmark for future research in forecasting within the RSC domain, particularly in parts remanufacturing.

However, Matsumoto and Komatsu (2015) also identifies critical limitations in the existing forecasting approaches, notably the absence of cause-effect mechanisms and the challenge of integrating external factors, such as weather conditions and product lifecycle stages, into the forecasting models. This gap highlights the need for advancing forecasting methodologies that can encompass a broader range of influencing factors, potentially enhancing accuracy and applicability in RSC forecasting.

The research by Syntetos et al. (2016) widens the discussion to the heuristic approaches employed in forecasting within CLSC, revealing a lack of scientific validity. This viewpoint opens up new opportunities for research into how hierarchical forecasting, demand variation, and supply forecasts might improve RL forecasting approaches, emphasising the importance of comprehensive models that take into account all aspects of projected output.

Clottey and Benton (2010) and Fleischmann et al. (2012) advocate for the adoption of quantitative forecasting methods that leverage historical data to predict future trends. These approaches, including time series models and associative models, could significantly improve the predictability and management efficiency of RL processes. However, there remains a notable reliance on management opinion for forecasting in some sectors, indicating an area for potential improvement through data-driven models.

Kumar et al. (2014) critically addresses forecasting challenges within the RSC, particularly focusing on the return rates of products. Through their research, they introduce an ANFIS tailored to navigate the complexities of forecasting in a multi-echelon, multi-period, and multi-product CLSC. This study highlights the paramount importance of accurately forecasting return product volumes as essential for enhancing profitability and efficiency in the RSC.

A notable gap outlined by Kumar et al. (2014) is the absence of integrative research that directly correlates the flows of returned products with the demand for new parts. This oversight signals a potential area for supply chain optimization that remains unexplored. Despite the promising utility of the ANFIS model in enhancing forecast accuracy and supporting strategic supply chain planning. Kumar et al. (2014) identify key limitations in their research, particularly the lack of comparative analysis with alternative forecasting methodologies such as neural networks and regression-based methods, as well as a scarcity of validation using real-world data.

2.2.1. Conclusion

This review of forecasting in the RSC highlights critical gaps that hinder optimal management, including a pronounced scarcity of data on return volumes (Toktay et al., 2003; Cui et al., 2020), a predominant focus on demand rather than supply forecasting of return defective parts (Thierry et al., 1995; Fleischmann, 2003), and a lack of methodological comparisons and empirical validations of forecasting models such as ANFIS against neural networks and regression methods (Kumar et al., 2014; Syntetos et al., 2016). Additionally, the literature underscores a significant oversight in not directly linking return flows with new part demands, a crucial aspect for enhancing supply chain efficiency (Kumar et al., 2014). Addressing these gaps is essential for advancing the accuracy and applicability of forecasting models within the RSC framework.

2.3. Failure Rate Stages & Reliability

This thesis focuses on predicting the return flow of defective machinery parts in the semiconductor industry during the initial assembly phase. Predictive analytics are necessary to estimate the quantity of defects based on historical data. Such defects are closely associated with the reliability and failure rates of machinery parts. Therefore, this section concentrates on examining literature related to reliability and failure rate predictions, exploring how these factors influence defect occurrences and return flows in semiconductor manufacturing.

According to (Lewis, 1996), quality and reliability, while distinct, are closely interconnected aspects crucial to product development. Quality is defined as the collection of features in a product or service that satisfies specific requirements at the time of manufacture. Reliability, on the other hand, relates to the consistent performance of a product throughout its expected lifespan under designated conditions, ensuring it operates without failure.

Lewis (1996) emphasizes that quality assessments are instantaneous, evaluating a product at a particular moment, whereas reliability considers the entire lifecycle of the product. This distinction is particularly important in industries like semiconductor manufacturing, where products are expected to function reliably over long periods and under diverse operational conditions. From this perspective, it is evident that quality issues are typically more pertinent to the initial stages of the manufacturing process.

Brombacher et al. (May 2005) discusses the nuanced facets of product reliability, noting that failures can be classified into physical and functional types. Physical failures occur due to material or component degradation, which may be instantaneous or progressive. Functional failures, however, arise when a product meets technical specifications but fails to align with user expectations or needs, reflecting a mismatch between design intentions and user requirements.

Three dimensions frame these reliability issues: adherence to specifications, statistical significance across various user scenarios, and the impact of time on failure rates (Brombacher et al., 2005). The ”bathtub curve” is a fundamental model illustrating these concepts, and is shown in Figure 2.4. It depicts an initial high failure rate (infant mortality), followed by a period of stable performance (useful life), and ending with increased failures due to aging (wear-out). Early failures (infant mortality) are often due to inherent deficiencies, such as missing components, substandard materials, or shipping damages. As these early issues are addressed, the failure rate declines, which is critical for establishing a baseline for reliable operation (Lewis, 1996).

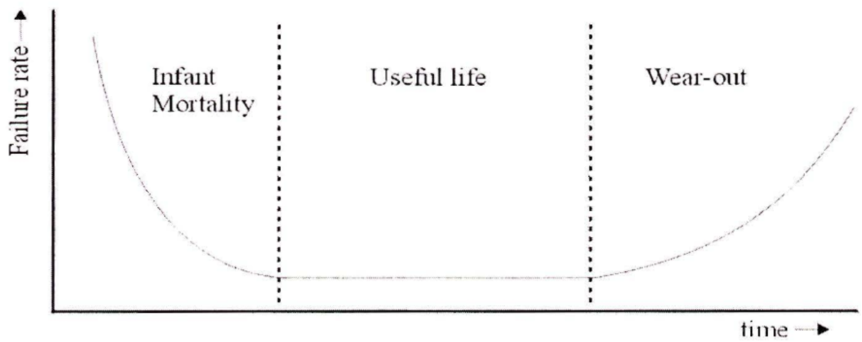


Figure 2.4: Bathtub Curve (Lewis, 1996)

Holcomb and North (Jan. 1985) introduced the concept of infant mortality in electronic components, underscoring a phase where failure rates are initially high but gradually decrease due to latent defects introduced during manufacturing, such as internal electrical flaws or contaminants. Roesch (Dec. 2012) expands on this by focusing on the ”extrinsic” phase of the bathtub curve, which is particularly relevant to the semiconductor industry. This phase is characterized by initially high failure rates resulting from manufacturing defects. This stage, often neglected in the context of semiconductors, especially in small batches or low-volume production, is critical for establishing benchmarks for subsequent reliability assessments. Roesch (Dec. 2012) highlights the criticality of early defect detection and resolution in boosting long-term reliability. These insights are fundamental for employing predictive analytics to effectively forecast defect rates during the assembly phase, thereby enhancing product reliability from the beginning.

Reliability forecasting in various industries often emphasizes stages beyond the initial ”infant mortality” phase of the product lifecycle, as represented by the bathtub curve. Notably, the Weibull distribution is extensively utilized in these later stages due to its effectiveness in modeling life data, which is crucial for predicting long-term product reliability (Lu, 1998). According to Abernethy (2006), the Weibull analysis is a commonly used method in engineering for determining component failures, leveraging its two primary parameters: the shape parameter, which influences the distribution’s skewness, and the scale parameter, which sets the threshold, thus determining the distribution’s location. However, the Weibull distribution is typically less relevant for the ”infant mortality” stage, where failures are more a result of manufacturing defects or early-use deviations rather than the wear-out mechanisms that dominate later life stages.

Focusing on the later stages of reliability enables organizations to plan maintenance more effectively, predict potential failures, and implement cost-effective replacement strategies. For example, Lee et al. (Nov. 2021) discusses various modeling approaches, including the parametric Weibull distribution, to forecast failures and enhance the reliability of automotive components. The use of the Weibull distribution in these contexts underscores its capability to provide accurate failure analyses even with minimal data, making it an invaluable tool in the arsenal of reliability engineers focused on extending the useful life of products well beyond their early usage phases.

Additionally, Lee et al. (Nov. 2021) explored different models, such as time-series methods like ARIMA and machine learning techniques including LSTM, support vector machines, and random forests. Lee et al. (Nov. 2021) indicates that these models often outperform the Weibull distribution, which is typically focused on later stages for reliability predictions. These alternative approaches are better suited for capturing the complexity and variability of early-stage failures, thereby providing more robust and accurate predictions for these critical initial phases.

Conclusion
This section explores the intersection of reliability and failure rates with the return of defective parts in semiconductor manufacturing, focusing on the initial assembly phase. Quality and reliability, while distinct, are interconnected and crucial for predicting defects (Lewis, 1996). The ”bathtub curve” model is central to understanding these aspects in this research project, highlighting the initial high failure rate due to infant mortality, followed by stable and wear-out phases (Lewis, 1996). While the Weibull distribution is effective in later life stages (Abernethy, 2006), it falls short during the infant mortality phase where defects are prevalent (Roesch, 2012). Alternative predictive models, such as ARIMA and machine learning techniques, are suggested to provide more accurate early-stage predictions (Lee et al., 2021). This thesis capitalizes on these insights by utilizing historical data in conjunction with predictive analytics to improve quality control and accurately forecast early-life failures in the assembly process.

2.4. Inventory Control & Classification

This section will first outline the inventory control of returned spare parts, followed by the classification of these parts.

2.4.1. Inventory Control

Forecasting the return of parts is essential and involves two pivotal decisions following assortment management, as identified by Driessen et al. (2014). The research illustrates these concepts through a diagram that provides an overview of processes and decisions in maintenance logistics control, as shown in Figure 2.5. Return parts forecasting is structured around two primary decisions:

1. **Classification of Parts:** Parts are classified either strategically or tactically based on their anticipated frequency of returns. This classification is generally an annual activity, aimed at categorizing parts to facilitate subsequent forecasting efforts.
2. **Forecasting Return Rates and Times:** Following classification, return rates and times are predicted on a more regular basis—typically monthly or quarterly. This step is crucial for accurately predicting the volume and timing of parts returns to the facility.

The relationship between parts return forecasting, inventory control, and demand forecasting is crucial. Accurate return forecasts aid in optimizing inventory by ensuring availability of necessary parts while minimizing excess stock. This is particularly critical for repairable parts, which might be restored to a ready-for-use state or scrapped if repair is unfeasible, significantly impacting operational efficiency and financial overheads (Driessen et al., 2014).

Furthermore, Driessen et al. (2014) highlights how return forecasting integrates into broader maintenance logistics frameworks. The data obtained from return forecasting informs decision-making in inventory control and aligns with demand forecasting to harmonize all aspects of spare parts logistics.

Below is a concise summary illustrating how parts return forecasting and inventory control connects directly and indirectly with various processes in maintenance logistics control (Driessen et al., 2014):

- **Assortment Management** (Direct): Influences decisions regarding which parts to include in the inventory based on their return rates and repairability.

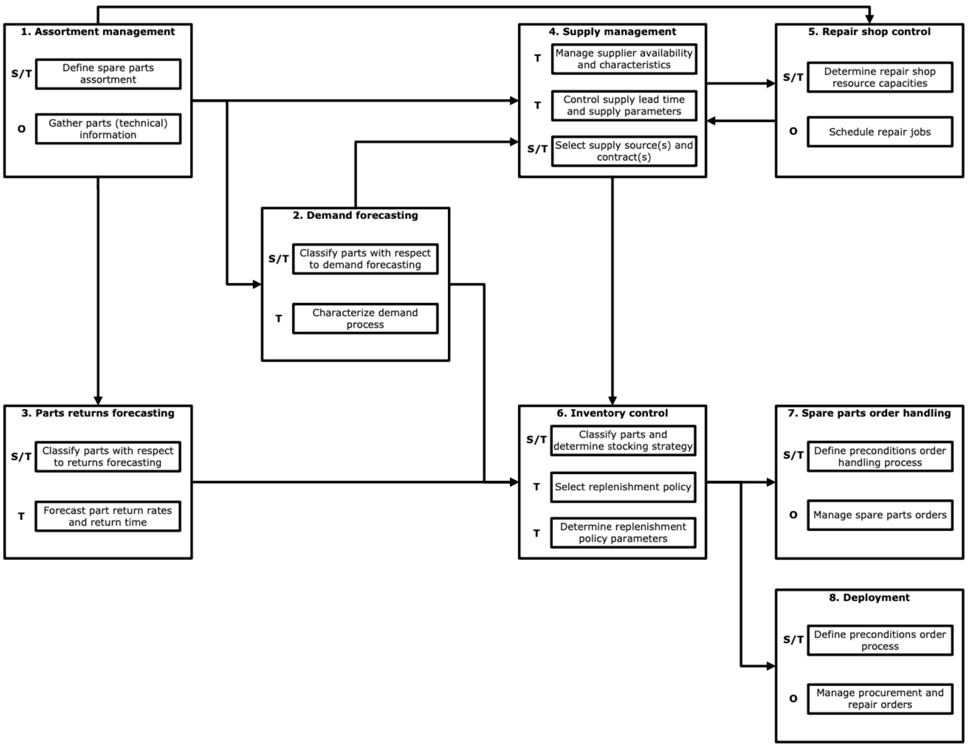


Figure 2.5: Diagram of Logistics Control Processes and Decisions (Driessen et al., 2014)

- **Demand Forecasting** (Direct): Enhances the accuracy of demand predictions by incorporating data on return rates and conditions of parts.
- **Supply Management** (Indirect): Affects supplier selection and contract negotiations by considering anticipated parts return rates.
- **Repair Shop Control** (Indirect): Guides scheduling and resource allocation in repair shops through forecasts of returned parts volumes and timing.
- **Deployment** (Indirect): Shapes replenishment policies and the management of procurement and repair orders by factoring in the flow and condition of returned parts. It also triggers adjustments in procurement and repair strategies based on discrepancies between forecasted and actual returns.

2.4.2. Classification of Spare Parts

As outlined by Driessen et al. (2014), the initial step in managing returns involves strategically classifying spare parts. Given the variety of spare parts in production and operations management, it is crucial to determine distinct characteristics of the stock, as detailed by Van Kampen et al. (June 2012). Since each part exhibits unique demand patterns, they often necessitate tailored forecasting approaches. However, it is impractical to develop individual models for each, thus grouping and applying specific forecasting models and control policies to each category is common practice Heinecke et al. (June 2013). Furthermore, classification aids in allocating managerial focus, particularly since not all returned items merit equal attention due to disparities in their value or volume (Molenaers et al., 2012). The classification methods outlined below are commonly utilized in spare parts inventory management and demand forecasting, which may also be applicable to forecasting the return of defective parts:

- **Demand Patterns:** The classification of demand patterns into smooth, intermittent, erratic, and lumpy categories by Syntetos and Boylan (2005) provides a systematic approach to forecasting and inventory management. This method is essential for understanding the frequency and variability of demand, which directly influences stock control strategies. The classification is based on two key metrics: the Average Demand Interval (ADI) and the Squared Coefficient of Variation (CV^2). ADI measures the average interval between successive demands, with lower values indicating more frequent demand. CV^2 quantifies the variability in demand size relative to the mean, with higher values indicating greater variability.

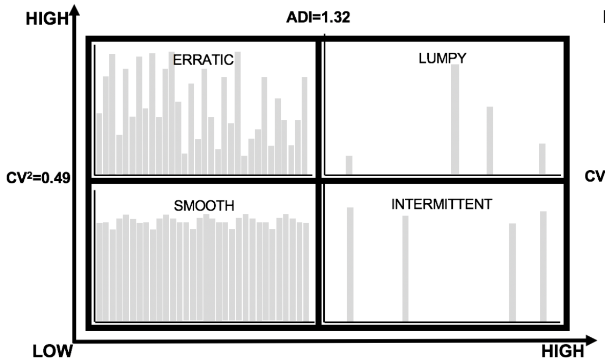


Figure 2.6: Demand Pattern Classification (Syntetos and Boylan, 2005)

As depicted in Figure 2.6, demand patterns are classified using cut-off values of ADI = 1.32 and CV^2 = 0.49. Smooth demand is characterized by ADI $<$ 1.32 and CV^2 $<$ 0.49, indicating regular and low variability demand. Intermittent demand, defined by ADI $\geq$ 1.32 and CV^2 $<$ 0.49, involves sporadic but predictable demands. Erratic demand, with ADI $<$ 1.32 and CV^2 $\geq$ 0.49, shows frequent and highly variable demand occurrences. Lumpy demand, identified by ADI $\geq$ 1.32 and CV^2 $\geq$ 0.49, features long periods without demand followed by high variability when demand occurs. The findings from Lamghari-Idrissi (2021) emphasize the practical implications of these categories in managing spare parts for ASML, particularly highlighting that lumpy demand patterns predominate (88.3%). Additionally, Bacchetti and Saccani (Dec. 2012) noted that spare parts typically encounter intermittent or lumpy demand patterns. Accurate classification of demand patterns is crucial for applying appropriate forecasting techniques and inventory policies, thereby optimizing cost and efficiency in supply chain operations.

- **ABC Analysis:** This technique categorizes items based on their importance and usage frequency. Typically, ”A” items account for 20% of the items but represent 80% of the total value, making them critical for inventory management. Syntetos et al. (Feb. 2009) used an extended version, ”ABCDEF”, for spare parts demand categorization, which refines this method. In this approach, ”A” items are those with more than one order in the last six months and an average demand quantity greater than six, indicating high-frequency demand. ”C” items have only one order in the last six months, while ”D”, ”E”, and ”F” items have no orders in the last 6, 12, or more than 12 months, respectively. This extended categorization aids in more accurate spare parts demand forecasting and enhances inventory management.
- **FSN Classification:** This method categorizes spare parts as fast-moving, slow-moving, or non-moving to optimize inventory management by identifying items for disposition and freeing up capital and space (Devarajan and Jayamohan, 2016). This method segments inventory into ”F” items (frequently ordered), ”S” items (lower turnover), and ”N” items (unused for an extended period), facilitating targeted management strategies (Stoll et al., 2015). However, practical applications, particularly using standard SAP time-series forecasting methods, often result in inaccuracies for slow-moving items, leading to excessive stock and heightened obsolescence risks (Syntetos et al., 2009).
- **VED Analysis:** VED analysis classifies items based on their criticality to the production process. ”V” items are vital and have a high impact, ”E” items are essential but with a lesser impact than ”V” items, and ”D” items are desirable but not critical (Jadhav and Jaybhaye, 2020).
- **HML Analysis:** This method categorizes inventory based on unit price. ”H” items are high-priced, ”M” items are medium-priced, and ”L” items are low-priced, facilitating financial management of inventory (Jadhav and Jaybhaye, 2020).
- **Analytic Hierarchy Process (AHP):** AHP is a multi-criteria decision-making method that ranks alternatives based on a hierarchy of criteria such as criticality, usage frequency, and lead time. This

method proves invaluable in classifying spare parts by incorporating both qualitative and quantitative factors into the inventory management strategy (Molenaers et al., 2012).

2.4.3. Conclusion

This section examines the crucial role of inventory control and classification of returned spare parts within RSC. Central to this process, as Driessen et al. (2014) outlines, is the accurate forecasting of return flows, which relies on strategic classification and regular monitoring of return rates and times. Such classification not only facilitates more precise predictions of parts returns but also optimizes inventory by aligning parts availability with operational demands, thereby reducing excess stock and increasing efficiency. The incorporation of forecast data enhances various logistical operations, from assortment management to repair shop control, shaping overall supply chain strategies. The various classification methods mentioned, including ABC, FSN, VED, HML, and AHP analysis, provide supply chain managers with valuable tools to prioritize and manage parts based on their value, usage frequency, and critical importance (Jadhav and Jaybhaye, 2020; Syntetos et al., 2009). This systematic approach to inventory and classification, utilized in spare parts inventory management, substantially aids in managing the complexities of RSC and may also be applicable to forecasting the return of defective parts.

2.5. Conclusion & Discussion

Each section of this literature review chapter concludes with key insights. This section synthesizes these insights to provide a comprehensive answer to the first sub-question of this research:

Research Question 1

What are the existing practices and challenges in managing the reverse supply chain for defective machinery parts?

In managing the RSC for defective machinery parts, there are several established practices and notable challenges. Practices include strategically integrating RSC within traditional supply chain frameworks to improve sustainability and efficiency. This involves adopting CLSC strategies that merge forward and RL, thereby extending product lifecycles and minimizing waste. Specifically tailored strategies are also developed for industries like automotive and electronics, which face unique challenges in managing return flows and aligning with sustainability goals.

The classification of demand patterns into smooth, intermittent, erratic, and lumpy categories provides a systematic approach to forecasting and inventory management, crucial for understanding demand frequency and variability. This method highlights that lumpy demand patterns predominate in managing spare parts for ASML, emphasizing the need for tailored forecasting techniques and inventory policies. Additionally, inventory management methods are essential, with advanced classification systems such as ABC, FSN, VED, HML, and AHP analysis aiding in the prioritization of items such as spare parts based on value and usage frequency. These systems optimize inventory levels by ensuring the availability of high-priority parts.

A major challenge is the scarcity of data on return volumes, which hampers the effectiveness of forecasting models crucial for planning return flows. The focus within the literature and practice predominantly remains on demand forecasting, with inadequate attention to supply-side forecasting for defective parts. This oversight restricts the development of precise predictive models crucial for RSC management. Furthermore, there is a lack of comprehensive comparisons and validations of different forecasting models, which stifles advancements in this area. Additionally, current models often fail to effectively link return flows with new part demands, which is critical for enhancing supply chain responsiveness and efficiency. Another critical gap is in the early stages of the product lifecycle (bathtub curve) where existing models like the Weibull distribution fall short in predicting defects accurately, necessitating alternative statistical approaches such as ARIMA and machine learning techniques for better early-stage defect predictions.

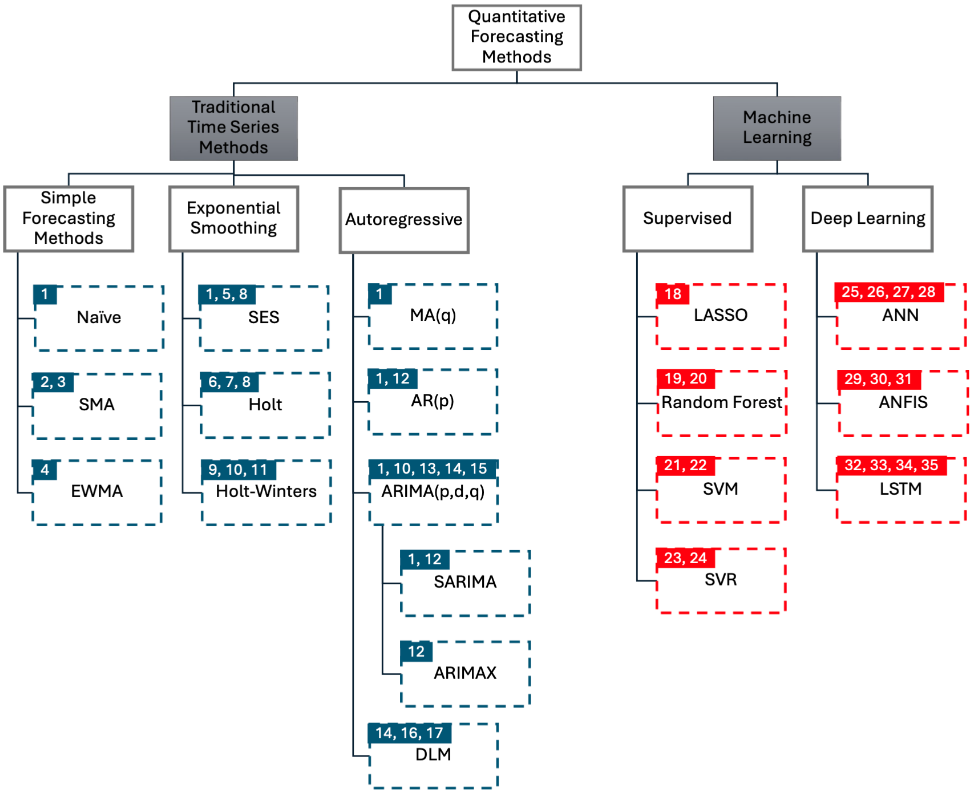
Addressing these challenges through focused research and methodological innovation is essential for enhancing the management capabilities within RSC, especially for defective machinery parts, contributing to more sustainable and efficient supply chain operations.

3

Literature Research: Forecasting Methods & Metrics

This chapter builds on the research gaps identified previously, focusing on various forecasting methods and models relevant to predicting return flows. An overview of all the forecasting methods and models is shown in Figure 3.1. This chapter evaluates their suitability for accurately forecasting the return of defective machinery parts and provides a detailed examination of the crucial variables and evaluation metrics integral to these models. By the end of this chapter, answers to the following sub-research question will be disccused:

2. *Which forecasting methods and models are available and suitable for predicting defective machinery parts in the return flow, and what are the crucial variables and evaluation metrics?*



1. Hyndman and Athanasopoulos, 2018	8. Krapp et al., 2013	15. Lee et al., 2021	22. Rostami-Tabar et al., 2013	29. Marx-Gomez et al., 2002
2. Johnston et al., 1999	9. Winters, 1960	16. Ma and Kim, 2016	23. Benti et al., 2023	30. Jin et al., 2013
3. Song and Li, 2008	10. Matsumoto and Komatsu, 2015	17. Geda and Kwong, 2018	24. Yuan et al., 2022	31. Kumar et al., 2014
4. Syntetos and Boylan, 2005	11. Chatfield and Yar, 1988	18. Cui et al., 2020	25. Wang et al., 2019	32. Hochreiter and Schmidhuber, 1997
5. Gardner, 1985	12. Box et al., 2015	19. Suthaharan, 2016	26. Elsaraiti and Merabet, 2022	33. Groenendijk et al., 2020
6. Holt, 1957	13. Turki et al., 2022	20. Sareminia and Amini, 2023	27. Dougherty, 1995	34. Ismail et al., 2018
7. Kumar and Yamaoka, 2007	14. Clottey et al., 2012	21. Zeng and Qiao, 2013	28. Temur and Bolat, 2017	35. Chandriah and Naraganahalli, 2021

Figure 3.1: Overview of Forecasting Methods & Models Discussed (Authors' own creation)

To comprehensively explore the forecasting methods and metrics related to this research, initial search terms included ”forecasting methods”, ”return flow prediction”, ”defective parts forecasting”, ”failure prediction”, ”supply chain” and ”reverse supply chain”. These search terms were specifically chosen to capture a broad spectrum of relevant literature and ensure a thorough examination of the various aspects of forecasting in the context of return flows and defective machinery parts.

To further refine the search and include seminal works and recent advancements, the snowballing technique was utilized. This involved reviewing the reference lists of key articles identified in the initial search to discover additional relevant studies.

3.1. Forecasting Methods & Models

Demand forecasting is a well-established discipline within SCM, as detailed by Syntetos et al. (2016). However, the area of return or supply forecasting, particularly in predicting early-stage defects within RSC, has not received comparable attention, as discussed in Sections 2.2 and 2.3. This section explores relevant forecasting methods and models to examine their objectives, functionalities, and potential modifications for accurately forecasting defects of machinery parts during the early failure stage.

3.1.1. Classification of Forecasting Methods

Forecasting methodologies are categorised from a broad perspective into qualitative and quantitative approaches. Qualitative methods are employed when there is a lack of historical data or the data is insufficient to establish patterns, often used in scenarios involving new products or rapidly changing market conditions where traditional data analysis is ineffectual (Caniato et al., 2011). Conversely, quantitative methods are preferred in stable environments where sufficient historical data can facilitate accurate forecasting. This research focuses on quantitative methods due to the availability of historical data related to defective machinery parts, enabling the precise quantification of future defects. These methods can be further divided into univariate and multivariate approaches. Univariate approaches consider a single variable, while multivariate approaches incorporate multiple variables to improve forecasting accuracy (Cerqueira et al., 2019).

Quantitative Forecasting Methods

Quantitative forecasting encompasses a spectrum of methodologies including causal models, time series analysis, and Machine Learning (ML) techniques. Recent advancements have blurred the lines between causal and ML methods, as ML increasingly incorporates causal inference to predict outcomes (Wang et al., 2021). This shift has led to a categorization of quantitative methods into two main groups: traditional time series models and ML models. The latter have gained significant traction over the past two decades due to their ability to handle complex datasets and provide robust predictive capabilities (Cerqueira et al., 2019; Rosienkiewicz et al., 2017).

Time Horizon

Forecasting horizons in SCM are classified into very short-term, short-term, mid-term, and long-term, each aligned with specific strategic goals (Makridakis et al., 1998). Very short-term forecasting (up to one month) addresses immediate operational decisions like daily production scheduling and inventory adjustments (Bowersox et al., 2002). Short-term forecasting (one to three months) focuses on tactical decisions such as monthly production plans and inventory restocking, incorporating market trends and seasonal factors (Chase et al., 2006). Mid-term forecasting (six to eighteen months) is vital for strategic decisions like capacity adjustments and long-term inventory strategies, integrating broader forecasts (Armstrong, 2001). This horizon is crucial for RSC management, particularly for planning remanufacturing and recovery processes (Guide and Wassenhove, 2009). Long-term forecasting (three to five years) supports strategic decisions like market expansion and new product development (Kotler and Armstrong, 2010). This research focuses on mid-term forecasting, as it aligns with the operational cycles of defect returns in RSC and possible capacity adjustments at the inventory level.

3.1.2. Traditional Time Series Models

Time series analysis employs historical data to forecast future events by identifying and leveraging data trends. This methodological framework involves constructing statistical models that encapsulate observed patterns from past data, facilitating predictions of future outcomes. Such models can range from univariate, focusing on a single variable, to multivariate approaches, which analyze simultaneous behavior across multiple variables. This technique is vital for understanding complex dynamics within datasets as it incorporates a variety of statistical methods to enhance prediction accuracy (Kalayci, 2003).

The core elements of time series analysis include the level, trend, seasonality, cycles, and irregular random fluctuations. The ’level’ sets the baseline value of the dataset, while the ’trend’ indicates the general directional movement, upward or downward. ’Seasonality’ accounts for consistent, predictable patterns that recur over specific intervals, and ’cycles’ represent longer-term economic or other fluctuations that do not follow a fixed periodicity (Hyndman and Athanasopoulos, 2018). ’Irregular random fluctuations’ encompass spontaneous, unpredictable variations. These components are integrated through time series decomposition, a critical process for distilling accurate forecasts from complex data sets (Silver et al., 2016).

The following subsections reviews traditional time series models relevant to SCM, with a specific focus on their use within RSC to forecast returns of defective machinery parts in the semiconductor industry. The analysis aims to identify which models best accommodate the distinctive characteristics of return flows, enhancing forecasting reliability.

Simple Forecasting Methods

The Naïve method stands as the simplest forecasting technique, relying exclusively on the most recent observation to predict future values. It assumes the immediate past observation is the optimal predictor for the upcoming period:

$$\hat{y}_{t+1|t} = y_t$$

where $\hat{y}_{t+1|t}$ is the forecast for the next period, and y_t represents the observed value in the current period. Celebrated for its simplicity, the Naïve method is particularly effective when the data shows minimal variation over time and has been proven robust across various economic and financial time series (Hyndman and Athanasopoulos, 2018). Despite its basic nature, it serves as a valuable benchmark against more sophisticated models, especially when the latest data point contains the most pertinent information for predictions (Paldino et al., 2021).

The Simple Moving Average (SMA) method calculates the arithmetic mean of the last n observations:

$$\text{SMA} = \frac{1}{n}\sum_{i=1}^{n} y_{t-i+1}$$

This method smooths out random fluctuations, aiding in the analysis of short to medium-term trends (Johnston et al., 1999). Although SMA is straightforward, its efficacy is limited in data lacking strong trends or seasonal variations (Song and Li, 2008). Within the RSC context, especially concerning stochastic return flows of defective parts, SMA may require adjustments to better address underlying trends typically overlooked by this model (Jacobs et al., 2014).

Building on SMA, the Exponentially Weighted Moving Average (EWMA) assigns progressively decreasing weights to older data, prioritizing more recent information:

$$S_t = \alpha y_{t-1} + (1-\alpha)S_{t-1}$$

Here, S_t represents the estimated value at time t, y_{t-1} the actual value from the previous period, and α the smoothing constant ($0 < \alpha < 1$). EWMA adapts more quickly to data changes, making it apt for scenarios where recent observations significantly influence future predictions (Syntetos and Boylan, 2005). Choosing the right α is crucial and should be tailored based on specific RSC data characteristics, such as repair times and failure rates, which critically affect forecasting accuracy.

SMA is well-suited for datasets with consistent return rates, whereas EWMA is more effective in situations where recent changes in return patterns significantly impact the accuracy of future forecasts. By adjusting the parameters of SMA and EWMA, specifically n for SMA and α for EWMA, these models can be better tailored to reflect the specific dynamics of defective part returns in the semiconductor industry. Such customization enhances the performance of these forecasting models.

Table 3.1 presents a summary of the relevant features, along with the comparative advantages and disadvantages of simple forecasting methods.

Table 3.1: Comparison of Simple Forecasting Methods

Method	Feature	Advantage	Disadvantage
Naïve	Uses the last observation as the forecast	Simple and robust for data with minimal variation	Ineffective for data with trends or seasonality
SMA	Average of the last n observations	Smooths fluctuations and aids in trend analysis	Limited effectiveness in non-trendy, non-seasonal data
EWMA	Weights recent data more heavily (α)	Adapts quickly to changes in data	Requires careful setting of the α parameter

Exponential Smoothing Methods

Exponential smoothing methods are essential for addressing various forecasting needs in time series analysis, particularly due to their flexibility in adjusting to data complexities. Simple Exponential Smoothing (SES) is the foundational method in this category, best suited for stable data without trends or seasonality:

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha)\hat{y}_{t|t-1}$$

where $\hat{y}_{t+1|t}$ is the forecast for the next period, y_t is the most recent observation, and α is the smoothing constant, modulating the impact of historical data (Hyndman and Athanasopoulos, 2018).

SES is frequently compared with the EWMA. While both methods emphasize recent observations, their applications differ significantly. EWMA primarily serves as a smoothing technique, laying the groundwork for further analysis or as part of more complex models. Conversely, SES is designed for direct forecasting, making it particularly valuable in environments requiring immediate and responsive predictions, such as flows within supply chains (Gardner, 1985).

For datasets with trends, double exponential smoothing, also known as Holt's Method, introduces a second parameter to handle trends (Holt, 1957). This approach proves essential in scenarios like vehicle consumption within CLSCs, where a linear trend is observable. Kumar and Yamaoka (2007) demonstrated the necessity of Holt's method in such environments, accommodating the linear trends detected in their data, which underscores the adaptability of exponential smoothing methods in practical applications.

When considering time series with both trends and seasonality, triple exponential smoothing, better known as Holt-Winters' Method, is preferred. This method integrates a seasonal adjustment to align forecasts with recurring cycles (Winters, 1960), crucial for predicting the flows subject to seasonal variations. The precision of Holt-Winters' method in handling such complex patterns was notably affirmed by Matsumoto and Komatsu (2015) in their study on remanufacturing environments, highlighting its low error margins and effectiveness in seasonal demand forecasting.

Additionally, challenges in applying Holt-Winters' method, such as the robustness of forecasts in varying conditions, were explored by Chatfield and Yar (1988), emphasizing the practical robustness of this method. Further, a study by Krapp et al. (2013) compared traditional forecasting methods, like Holt's method, to Bayesian estimation techniques. While Holt's method showed lower accuracy in return forecasts, it still achieved an acceptable level of performance.

The choice among these methods is influenced by the specific characteristics of the data, making their selection crucial for achieving accurate and reliable forecasts in RSC. SES is optimal for stable data, Holt's method is advantageous for datasets with linear trends, and Holt-Winters' excels in scenarios with compounded trends and seasonalities.

Table 3.2 presents a summary of the relevant features, along with the comparative advantages and disadvantages of exponential smoothing methods.

Autoregressive Models

Auto-Regressive (AR) models are fundamental in time series forecasting, especially in scenarios with significant autocorrelation. Formally, an AR model of order p (denoted as AR(p)) is represented by the equation:

Table 3.2: Comparison of Exponential Smoothing Methods

Method	Feature	Advantage	Disadvantage
SES	Single smoothing constant (α)	Direct and responsive forecasting for stable data	Limited by lack of trend or seasonality adaptation
Holt	Introduces a trend component	Effective for linear trend data	May overlook non-linear trends or seasonality
Holt-Winters	Incorporates both trend and seasonal adjustments	Ideal for handling complex patterns with trends and seasonality	More intricate, requiring precise calibration of parameters

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

where y_t indicates the current value, c is a constant, ϕ coefficients signify the impact of previous values, and ε_t denotes white noise (Hyndman and Athanasopoulos, 2018).

In the context of return forecasting in the semiconductor industry, AR models can help predict return flows by analyzing patterns in previous returns, which is crucial for managing RL effectively (Box et al., 2015).

The Integration component (I), involves differencing the series d times to render it stationary—a necessary precondition for the effective application of ARIMA models, which combine AR and Moving Average (MA) components (Hyndman and Athanasopoulos, 2018):

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

This MA component accounts for lagged forecast errors, aiding in smoothing out noise and enhancing forecast clarity.

ARIMA models, by integrating AR, I, and MA components, address a wide array of time series data, including those with non-stationarity and seasonal fluctuations (Box et al., 2015). They are extensively used for forecasting, adaptable through the Box-Jenkins methodology for precise model fitting and parameter estimation (Hyndman and Athanasopoulos, 2018).

Furthermore, the use of ARIMA models necessitates careful consideration of their constraints, notably the requirement for data stationarity and the model's challenges with handling high-frequency seasonal variations. Such issues often necessitate significant modifications, such as the adoption of Seasonal ARIMA (SARIMA) models, to accommodate seasonal patterns effectively (Box et al., 2015).

However, traditional ARIMA models sometimes fail to capture complex dependencies such as those between past sales and future returns, which can be crucial in predicting returns of defective parts. This limitation is highlighted in studies by Clottey et al. (2012), who suggest that incorporating bivariate models or extensions such as ARIMAX could enhance forecasting accuracy by integrating external variables like warranty claims or failure rates. ARIMAX is essentially an ARIMA model that includes one or more exogenous variables, directly affecting the forecast.

In the study discussed earlier, Matsumoto and Komatsu (2015) not only investigates the effectiveness of the Holt-Winters model but also compares it to the ARIMA model for demand forecasting in the context of auto spare parts remanufacturing. The comparison between these two models provides a comprehensive overview of how each model handles the inherent complexities of forecasting in a remanufacturing environment, particularly in terms of capturing seasonal and periodic variations in sales data. This analysis offers valuable insights into the suitability of these models for enhancing the accuracy of forecasts in production planning specific to remanufacturing scenarios.

In their study on forecasting spare part extractions in a CLSC, Turki et al. (2022) assess the efficacy of ARIMA models, highlighting their precision in structured forecasting scenarios. Despite ARIMA's strengths in parameter customization and accuracy for higher volume parts, its limitations include an inability to process non-linear trends and seasonal fluctuations without modifications. This necessitates the integration of ARIMA with other models to enhance forecasting reliability in dynamic supply chain environments (Turki et al., 2022).

In the study by Lee et al. (Nov. 2021), an ARIMA model was utilized to forecast the number of failures in automobile parts based on time series data from the first six months of usage. The data derived from two specific vehicle models. The ARIMA model parameters tested across a range of values *(p = 1 to 4; d = 0 to 2; q = 1 to 4)* to pinpoint the combination that best modeled the underlying patterns of failures while optimizing for simplicity, as indicated by the lowest Akaike Information Criterion (AIC).

Once the optimal parameters were determined for each part, the ARIMA model was employed to predict future failures from the seventh to the sixtieth month. This predictive modeling aimed to enable proactive quality assurance measures and reduce warranty service expenses by anticipating potential failures. Despite these efforts, the predictive accuracy of the ARIMA model fell short when compared to more advanced machine learning approaches, such as deep learning, which proved more effective in forecasting failures and estimating long-term reliability more precisely (Lee et al., 2021).

Following the discussion on ARIMA models, Distributed Lag Models (DLM) provide another robust approach for modeling the relationship between past sales and future returns, crucial in industries with significant return rates like electronics and automotive. DLMs conceptualize returns (m_t) as a cumulative function of sales over previous periods, expressed by:

$$m_t = \sum_{k=1}^{T} \beta_k n_{t-k} + \epsilon_t$$

where n_{t-k} represents sales, β_k are the lag coefficients depicting the impact of these sales on returns, and ϵ_t is an error term, typically modeled as white noise (Clottey et al., 2012).

The strength of DLM lies in its adaptability; parameters (β) can be updated as new data emerges, enhancing forecast accuracy over time (Clottey et al., 2012). This feature is beneficial in dynamic market environments where product life cycles and sales patterns are subject to change. Parameters are generally estimated using ordinary least squares when lags are manageable, but more complex scenarios might require advanced techniques for greater accuracy (Geda and Kwong, 2018).

DLMs are shown to be most effective when the sales-returns relationship is strong and stable. However, in environments where this relationship is weaker or variable, mixed models that incorporate elements of ARIMA and regression could provide better forecasting accuracy (Ma and Kim, 2016).

A challenge with DLMs is determining the appropriate lag length and ensuring accurate parameter estimation, as errors in these areas can significantly degrade the model's effectiveness (Geda and Kwong, 2018). Despite these challenges, DLMs offer a robust framework for forecasting returns, providing valuable insights for managing RL in sectors with significant return volumes.

Table 3.3 presents a summary of the relevant features, along with the comparative advantages and disadvantages of auto-regressive methods.

3.1.3. Machine Learning Methods

This subsection reviews ML methods, as grouped by (Benti et al., 2023) into four categories: supervised learning, unsupervised learning, reinforcement learning, and deep learning. The focus will be on supervised and deep learning due to their relevance to the labeled data used in this study, which is essential for forecasting applications.

Supervised learning involves training models on datasets with labeled input-output pairs, optimizing for accuracy in predicting specific outcomes. This makes it particularly suited for regression and classification tasks within structured data environments. Deep learning, characterized by its use of multi-layered neural networks, is adept at processing and learning from large volumes of complex data, thereby enhancing the capability and precision of forecasting models (Benti et al., 2023).

For this review, unsupervised and reinforcement learning will not be covered due to their less direct applicability to the structured, labeled datasets at hand. The focus will remain on how supervised learning and deep learning can be optimally applied to address forecasting challenges, leveraging their robust capabilities in handling labeled and intricate data patterns effectively.

Table 3.3: Comparison of Autoregressive and Related Forecasting Methods

Method	Feature	Advantage	Disadvantage
AR(p)	Utilizes past values $(\phi_1, \phi_2, ..., \phi_p)$	Effective for data with autocorrelation, captures dynamics over multiple past periods	Requires stationarity; not suitable for non-linear trends
MA(q)	Uses past forecast errors $(\theta_1, \theta_2, ..., \theta_q)$	Good for smoothing out noise and short-term fluctuations	Does not account for long-term trends or seasonality
ARIMA	Combines AR, I (differencing), and MA components	Flexible, handles non-stationary data including trends	More complex to configure and calibrate, requires careful differencing
SARIMA	Extends ARIMA with seasonal adjustments	Specifically targets seasonal data variations, offering precise adjustments	More parameters to estimate, which can increase model complexity and fitting challenges
ARIMAX	Extends ARIMA to include exogenous variables	Enhances forecasting accuracy by incorporating external influences	Increases model complexity, requiring more data for effective estimation
DLM	Models returns as a function of distributed lag effects from past sales	Captures delayed impact of sales on returns, adaptable to new data	Requires careful selection of lag length and parameter estimation

Supervised Learning
The LASSO model, renowned for its dual functionality in demand forecasting—variable selection and regularization—effectively minimizes overfitting while identifying pertinent predictors from expansive datasets. These datasets may incorporate variables like usage patterns, production batch specifics, and historical failure rates (Cui et al., 2020). The mathematical formulation of LASSO is:

$$\text{Minimize: } \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

Here, y_i represents the return volumes, x_{ij} the predictors, β_j the coefficients, λ the regularization parameter, and p the number of predictors. This model's capacity to streamline complex datasets by excluding extraneous variables makes it highly effective in SCM, improving both interpretability and manageability.

Cui et al. (2020) acknowledges the model's prowess in forecasting product returns, yet points out its shortcomings in specifically identifying defects as the cause of returns. In large datasets characteristic of RSC scenarios, where multiple factors affect return probabilities, LASSO's attribute of performing variable selection and regularization concurrently proves invaluable. Cui et al. (2020) illustrates this with a case involving over 331,000 products, where LASSO successfully isolated critical predictors such as sales, product types, and historical returns. This pinpointing of relevant variables is essential for precise return volume forecasts, which in turn, are crucial for optimizing inventory and minimizing waste in sectors where returns play a essential role.

The Random Forest algorithm excels in addressing the nonlinear and complex relationships characteristic of RSC data through an ensemble of decision trees. It improves prediction accuracy by employing bootstrap sampling and random feature selection to mitigate overfitting (Suthaharan, 2016). The method is particularly advantageous for its ability to estimate variable importance, which could be crucial for pinpointing key predictors of defective machinery parts.

The operational mechanism of Random Forest includes (Suthaharan, 2016):

- Bootstrap Sampling: Creating diverse training subsets from the original dataset, allowing each tree to learn from different samples.
- Feature Randomness: During tree splits, considering a random subset of features to enhance model diversity and prevent overfitting.
- Aggregation of Tree Results: Averaging the predictions from all trees to produce a final, more stable output.

The capability of Random Forest to rank variables by importance is demonstrated through evaluation metrics such as Mean Decrease Accuracy, assessing how predictor shuffling affects model accuracy. This feature is vital for identifying critical factors influencing defect occurrences and return likelihood. Studies, such as those conducted by Sareminia and Amini (2023), demonstrate that Random Forest performs well in predicting spare parts demand, although it requires a substantial amount of data points to be effective.

Support Vector Machine (SVM) are well-regarded for their dual capability in classification and forecasting, making them particularly versatile across various data contexts. Traditionally employed to maximize the margin between classes of data points, SVMs are adept at ensuring precise categorization, a critical feature in settings requiring accurate predictions. For example, a study by Zeng and Qiao (2013) demonstrated that a 2D least-square SVM could outperform traditional forecasting models like AR models and radial basis function neural networks in short-term prediction. This capability to accurately predict outcomes is enhanced by SVMs' utilization of a wide range of input features, including historical and environmental data. Moreover, their adaptability was further evidenced in a different study by Rostami-Tabar et al. (2013), where SVMs were used to forecast demand for oil industry spare parts using a piecewise linearization approach. This method proved effective in managing uncertainty, showcasing SVMs' value in complex supply chain scenarios where conventional models may falter.

Support Vector Regression (SVR) is an advanced ML technique used for predicting continuous outcomes in regression analysis. SVR operates by mapping data into a higher-dimensional space using the kernel trick, a method that allows the model to handle non-linear relationships effectively (Benti et al., 2023). This mapping facilitates the fitting of the best possible hyperplane that minimizes the error margin between the actual and predicted values. One of the critical features of SVR is its flexibility in choosing kernel functions and setting regularization parameters, which are vital for optimizing the model's performance in diverse forecasting environments. This capability makes SVR particularly useful in sectors where the relationships between variables are complex and non-linear.

A notable application of SVR in the renewable energy sector is demonstrated by Yuan et al. (2022), who employed an optimized SVR model to predict wind power. Their study highlighted the model's superior performance over traditional methods, particularly in terms of handling seasonal variability and improving prediction accuracy, underlining SVR's robustness and adaptability.

Table 3.4 presents a summary of the relevant features, along with the comparative advantages and disadvantages of supervised learning methods.

Table 3.4: Comparison of Supervised Learning Methods

Method	**Feature**	**Advantage**	**Disadvantage**
LASSO	Regularization and variable selection	Efficiently handles large datasets with many predictors, reducing overfitting and enhancing model interpretability.	May fail to capture complex, non-linear relationships.
Random	Ensemble of decision trees, bootstrap sampling, feature randomness	Excellent at handling non-linear data; provides insights on variable importance and robustness against overfitting.	Computationally intensive; model performance heavily depends on correct parameter tuning.
SVM	Classification and regression capabilities using optimal hyperplanes	Versatile and effective in high-dimensional spaces; good for both classification and regression tasks.	Computationally intensive; performance heavily reliant on appropriate kernel choice.
SVR	Regression with kernel trick to handle non-linear data	Effective in forecasting scenarios with non-linear relationships; robust across various applications.	Sensitive to kernel and regularization settings; may require extensive parameter tuning.

Deep Learning
Deep learning models represent an advanced branch of machine learning, characterized by their use of Artificial Neural Networks (ANN) with extensive layering, suitably named "deep". These layers can range from just a few to several thousand (Wang et al., 2019). The primary function of these models is to identify

patterns and connections in data by dynamically adjusting the neural connections to reduce errors between predicted and actual outputs (Elsaraiti and Merabet, 2022). Structurally, ANN are designed to mimic the human brain, beginning with an input layer that receives data, usually in the form of numerical vectors, and concluding with an output layer that delivers results, which could be classifications or predictions (Dougherty, 1995).

In the specific context of managing RL for Waste of Electrical and Electronic Equipment (WEEE), Temur and Bolat (2017) have effectively utilized ANN to forecast the volumes of product returns accurately. Their study highlights the adaptability of ANN in processing complex data, a critical aspect for optimizing reverse logistics frameworks. The researchers tested various configurations of ANN, varying in the number of hidden layers and neurons, to discover the most effective structure for accurate predictions. They found that a model with two hidden layers, each containing four neurons, achieved the lowest mean squared error and provided highly reliable forecasts. This model was particularly effective in predicting return volumes for new collection points anticipated under future regulatory frameworks. The ability of ANN to deliver dependable forecasts is essential for the strategic planning and implementation of efficient RL operations, ultimately improving the management systems for WEEE (Temur and Bolat, 2017).

Building on these advancements, Marx-Gomez et al. (2002) introduced a neuro-fuzzy system that further refines the forecasting of product returns, especially for scrapped electronic products. Unlike traditional methods, their neuro-fuzzy approach utilizes both fuzzy logic and neural networks to handle the inherent uncertainties in return data, such as timing and quantity. This system incorporates expert knowledge through fuzzy rules and uses a sophisticated inference mechanism to predict returns more accurately. The success of their model in a case study with photocopiers demonstrates its potential to enhance prediction capabilities in RL, particularly under varying conditions dictated by real-world complexities (Marx-Gomez et al., 2002).

The ANFIS model uniquely blends the adaptive learning capabilities of neural networks with the heuristic reasoning of fuzzy logic, making it particularly adept at handling complex and nonlinear systems. According to Jin et al. (Aug. 2013), ANFIS is structured into five distinct layers: input fuzzification, rule application, normalization of firing strengths, defuzzification, and output summation. Figure 3.2 provides an illustration of ANFIS architecture by Jin et al. (Aug. 2013). This architecture allows ANFIS to fine-tune membership functions and if-then rules directly from the data, significantly enhancing its precision in time series forecasting. Employing a hybrid learning algorithm that integrates gradient descent and least squares methods, ANFIS adjusts its model parameters for optimal accuracy. While offering substantial benefits in modeling intricate data relationships, challenges such as potential overfitting and high computational demands during training are notable drawbacks, particularly when the data is noisy or the ruleset extensive (Jin et al., 2013).

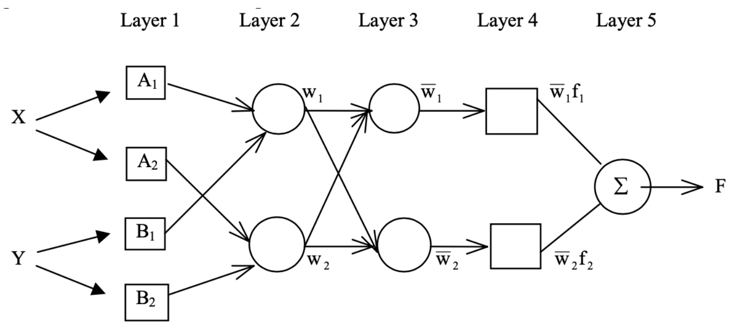


Figure 3.2: An ANFIS architecture by Jin et al. (Aug. 2013)

Building on this foundation, Kumar et al. (2014) applied ANFIS to forecast return products within an integrated forward and reverse supply chain, addressing the uncertainties prevalent in the volume of product returns. The approach leverages the neural network capabilities and fuzzy logic of ANFIS to significantly improve forecasting accuracy under conditions of incomplete data and complex patterns, which traditional forecasting methods often fail to address effectively. The methodology encompasses a two-phase process: initially, ANFIS forecasts product return rates, adapting to the dynamics of multi-period, multi-product, and multi-echelon closed-loop supply chain settings, factoring in historical sales data, consumer return tendencies, and existing incentives. Subsequently, the forecasted data inform the optimization of the supply chain design to ensure robustness and efficiency (Kumar et al., 2014).

The evaluation of the ANFIS model across 25 periods highlighted its reliability and high accuracy in forecasting, promising substantial applicability for real-world CLSC challenges. The evaluation metrics used demonstrated a close match between the forecasted and actual return volumes, affirming ANFIS's effectiveness in managing the uncertainties associated with return product forecasting (Kumar et al., 2014).

LSTM, developed by Hochreiter and Schmidhuber (1997), are an advanced form of Recurrent Neural Networks (RNN) engineered to overcome the limitations posed by the vanishing gradient problem typical in traditional RNNs. These networks feature specialized memory cells regulated by three distinct gates: the input gate (i_t), the forget gate (f_t), and the output gate (o_t). These gates selectively retain or discard information, facilitating effective long-term sequence processing and allowing for robust error backpropagation across multiple layers, thus enhancing the network's capability to learn from long sequences (Hochreiter and Schmidhuber, 1997). Subsequent studies, such as those by Groenendijk et al. (2020), have underscored the ability of LSTM to propagate errors efficiently and bolster learning capabilities across complex datasets.

Figure 3.3 illustrates an LSTM cell in detail, highlighting its core components: the input X_t, the previous hidden state h_{t-1}, and the previous cell state C_{t-1} (Ismail et al., 2018). The operational mechanism within the cell includes the modulation gate ($\tilde{C}_t$) alongside the aforementioned gates. The forget gate determines parts of the cell state to retain or discard, the input and modulation gates update the cell state with new information, and the output gate determines the final output, which, after passing through a $tanh$ function, becomes the new hidden state h_t. This intricate gating system ensures critical data is preserved over extended periods, crucial for tasks requiring medium and long-term data retention (Ismail et al., 2018).

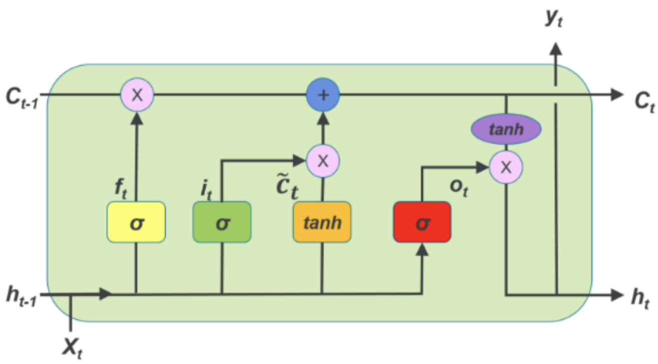


Figure 3.3: An LSTM cell (Ismail et al., 2018)

Chandriah and Naraganahalli (2021) demonstrated the effectiveness of LSTM in predicting automobile spare parts demand. Their model, trained on historical data and installed base information, achieved superior accuracy compared to traditional methods, even for parts with delayed installation. This approach has the potential to improve reverse logistics and maintenance scheduling.

Lee et al. (Nov. 2021) showcased the use of LSTM for predicting automobile part failures using warranty claim data. LSTMs' ability to handle temporal dependencies in sequential data is crucial for capturing the irregular patterns indicative of part failures. Their "many-to-many" LSTM architecture outperformed other models in accuracy metrics, making them possibly suitable for industries like semiconductor manufacturing to predict defective machinery parts.

Table 3.5 presents a summary of the relevant features, along with the comparative advantages and disadvantages of deep learning methods.

3.2. Evaluation Metrics

Forecasting the return flow of defective parts in the requires the application of sophisticated models whose accuracy needs to be critically assessed. This review dissects the most relevant evaluation metrics used in the literature, providing a clear understanding of their formulas, applications, and theoretical underpinnings.

Table 3.5: Comparison of Deep Learning Methods

Method	Feature	Advantage	Disadvantage
ANN	Multi-layered artificial neural network	Capable of processing complex and non-linear data patterns, good for dynamic data.	Requires large datasets for training; prone to overfitting without careful regularization.
ANFIS	Adaptive learning capabilities integrating fuzzy logic	High precision in handling non-linear systems; effectively tunes membership functions and rules from data.	Complex system setup; potential for overfitting; computationally intensive, especially with extensive rule sets.
LSTM	Recurrent neural network with long-term dependency handling	Excels in managing data where temporal sequences and past information are crucial, suitable for forecasting part failures.	High complexity leading to challenging model tuning and longer training times.

3.2.1. Scale-Dependent Measures

Scale-dependent measures are vital when forecast and actuals are on the same scale, allowing for direct comparison of model performance.

Mean Squared Error (MSE)

The MSE quantifies the average squared deviation between forecasts and actual observations, serving as a foundational measure of forecast error magnitude. Despite its known sensitivity to large errors, it is widely utilized in quantitative finance and forecasting due to its clear interpretative advantages in scenarios where large errors are particularly detrimental (Hyndman and Koehler, 2006).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.1}$$

Root Mean Squared Error (RMSE)

The RMSE is preferred in various forecasting scenarios due to its capacity to penalize larger errors more significantly, thus helping to highlight models that may underperform during critical operational situations (Chai and Draxler, 2014).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.2}$$

Mean Absolute Error (MAE)

The MAE, is frequently selected for its straightforwardness and robustness against outliers. This makes it an indispensable metric for comparing either a single time series or multiple series that are measured on the same scale. Its simplicity in calculation and interpretation renders it highly suitable for practical applications within SCM (Hyndman and Athanasopoulos, 2018). The MAE is also commonly referred to as the Mean Absolute Deviation (MAD), where the ”D” stands for ”deviation”, utilizes absolute values to calculate deviations, ensuring that negative and positive errors do not cancel each other out, thus providing a more accurate measure of forecast accuracy (Hyndman and Koehler, 2006).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.3}$$

3.2.2. Scale-Independent Metrics

Scale-independent metrics are crucial for comparing models applied to different scales of data.

Mean Absolute Percentage Error (MAPE)
MAPE offers a way to express forecast error as a percentage, making it a preferred choice for cross-scale model performance comparison (Armstrong and Collopy, 1992). It allows for straightforward interpretation, though it is not without its issues, particularly when datasets include zeros or extreme values.

$$MAPE = 100\% \times \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{3.4}$$

Symmetric Mean Absolute Percentage Error (SMAPE)
SMAPE addresses some of the known limitations of MAPE by accounting for both under- and over-forecasts symmetrically, thus avoiding distortion from extreme values (Makridakis et al., 2020).

$$SMAPE = 100\% \times \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{\frac{|y_i| + |\hat{y}_i|}{2}} \right| \tag{3.5}$$

Modified Mean Absolute Percentage Error (mMAPE)
The mMAPE is designed to address the limitations of standard MAPE, particularly in datasets with zero or near-zero values. Standard MAPE can yield undefined or extremely high values in such cases. The mMAPE enhances robustness and interpretability by capping the error and ensuring the denominator is never zero. The term $(1 + |y_i|)$ in the denominator ensures this condition, where y_i is the actual value and $\hat{y}_i$ is the forecasted value (Turki et al., 2022).

$$mMAPE = 100\% \times \frac{1}{n} \sum_{i=1}^{n} \left(\frac{|\hat{y}_i - y_i|}{1 + |y_i|} \right) \tag{3.6}$$

3.2.3. Synthesis

This synthesis consolidates findings from multiple earlier discussed studies focusing on the accuracy of the forecasting models.

The study by Krapp et al. (2013) evaluated the efficacy of SES and Holt's method using MSE, MAPE, and MAD. It was found that Holt's method produced higher MSE and MAD values, indicating larger errors compared to the Bayesian estimation approach. Despite this, Holt's method showed better accuracy than SES.

In research conducted by Matsumoto and Komatsu (2015), the ARIMA and Holt-Winters models were assessed using data from auto parts remanufacturing. ARIMA demonstrated superior performance, providing more accurate predictions over both short and long terms using SMAPE and consistently outperforming Holt-Winters in MAPE.

Clottey et al. (2012) applied MAPE to evaluate the performance of the DLM. Meanwhile, Lee et al. (Nov. 2021) used MSE and MAE to measure the prediction performance of models like ARIMA, SVM, and LSTM over various periods, establishing a ranking based on these metrics to compare their effectiveness.

In another study, Temur and Bolat (2017) employed MSE to gauge the performance of ANN models, and Kumar et al. (2014) assessed the ANFIS model using the coefficient of determination (R^2) along with RMSE, MAE, and MAPE. Additionally, Chandriah and Naraganahalli (2021) used MSE to evaluate LSTM models.

The literature review reveals that among the scale-dependent metrics, RMSE and MAE are preferred for their capacity to highlight significant forecasting errors and their ease of interpretation (Chai and Draxler, 2014; Hyndman and Athanasopoulos, 2018). MAPE, despite its popularity, requires careful application, especially in data sets with zero values, to avoid misinterpretation (Armstrong and Collopy, 1992). SMAPE's symmetric consideration of errors offers an improved alternative, particularly in data sets prone to outliers (Makridakis et al., 2020). The mMAPE is designed to address the limitations of standard MAPE, particularly in datasets with zero or near-zero values, and has been utilized for evaluating ARIMA models among others (Turki et al., 2022).

Table 3.6 provides a summary of the evaluation metrics applied across various studies to assess different relevant forecasting models for this research.

Table 3.6: Evaluation of forecasting models across different metrics, categorized by cited studies.

Citation	MSE	RMSE	MAE	MAPE	SMAPE	mMAPE
(Krapp et al., 2013)	SES	-	-	SES	-	-
	Holt	-	-	Holt	-	-
(Matsumoto and Komatsu, 2015)	-	-	-	ARIMA,	ARIMA,	-
				Holt-Winters	Holt-Winters	-
(Clottey et al., 2012)	-	-	-	DLM	-	-
(Lee et al., 2021)	ARIMA,	-	ARIMA,	-	-	-
	SVM,		SVM,			
	LSTM		LSTM			
(Temur and Bolat, 2017)	ANN	-	-	-	-	-
(Kumar et al., 2014)	-	ANFIS	ANFIS	ANFIS	-	-
(Chandriah and Naraganahalli, 2021)	LSTM	-	-	-	-	-
(Turki et al., 2022)	-	-	-	-	-	ARIMA

3.3. Conclusion & Discussion

Each section of this literature review chapter concludes with key insights. This section synthesizes these insights to provide a comprehensive answer to the second sub-question of this research:

Research Question 2

Which forecasting methods and models are available and suitable for predicting defective machinery parts in the return flow, and what are the crucial variables and evaluation metrics?

To address the urgent need for robust forecasting within the RSC of the semiconductor industry, this discussion synthesizes insights from a comprehensive literature review and a detailed matrix of forecasting models discussed (Figure 3.4). The primary objective is to identify models that can accurately predict the return flow of defective machinery parts, thereby mitigating challenges such as overstocked inventories and inefficiencies in managing these returns.

Traditional forecasting methods, such as SMA, EWMA and SES provide foundational approaches but are generally limited to simpler scenarios. These methods often rely on sales data to forecast future demand, which is more typical in B2C contexts and less suited to the complexities of defective part returns in the semiconductor industry. Among less complex models, Holt-Winters stands out for its ability to handle variations from return flow more effectively, offering a balance between simplicity and adaptability to data fluctuations.

Advanced models like ARIMA, ARIMAX, and LSTM have shown significant potential in addressing the unique challenges associated with spare parts and failure rates in RSC. ARIMA models are adaptable to non-stationary data, a common characteristic of return flow and spare parts demand. ARIMAX extends ARIMA by integrating external variables, such as repair rates and failure statistics, which are crucial for predicting return flows. These models adjust well to data irregularities inherent in manufacturing processes, effectively forecasting defect quantities.

LSTM models excel in managing medium and long-term dependencies and recognizing complex patterns. They are particularly suited for scenarios where past events significantly influence future outcomes, such as parts failing during initial assembly stages in ASML's operations. LSTM models have demonstrated superior performance compared to ARIMA, SVM, Random Forest, and ANN in scenarios involving complex dependencies and sparse data, which are typical in forecasting spare parts demand.

However, models like ANFIS and LASSO, although valuable for specific tasks such as planning efficiencies or analyzing return causes, are less preferred due to their complexity or lack of direct focus on quantitative forecasting needs, particularly in operational environments characterized by high defect variability.

Because of the different data variations of defective machinery parts there is need to investigate different models which can deal with sparse data, but for some more and some less complex so a choice would be favorable to test different complexities.

The discussion highlights the crucial variables for effective forecasting in the supply chain. Key variables include historical defect rates, which help establish baseline trends; production data, which may indicate potential manufacturing flaws; and sales data, although in this context, sales data is less relevant as it pertains more to customer returns rather than returns from early-stage factory processes. By analyzing these variables, forecasting models can be precisely adjusted to reflect the real-world complexities of the semiconductor manufacturing process.

The literature review highlights the importance of selecting appropriate evaluation metrics for forecasting return flows in the semiconductor industry. MAE and MSE are particularly valued for their ability to emphasize significant forecasting errors and offer clear interpretations, making them suitable for complex supply chain environments. MAPE, though widely used, requires careful consideration in datasets with zero values, whereas mMAPE offers a more robust alternative. SMAPE provides a balanced error assessment, advantageous in datasets with outliers. This synthesis underscores the need for tailored metrics to ensure accurate and reliable forecasting.

		General & Sector Specific						Forecasting Characteristics								
Forecasting Models	References	Logistics / SC Usage	Return Flow	Failure / Defect of Parts	Spare Parts Demand Pattern	Sales Input Variable	Complex Patterns	Time Series Based	Multivariate	Non-Stationary	Trend	Seasonality	Time Horizon (**S**hort, **M**edium, **L**ong)	Complexity	Data Volume Required	Performence for Large Variation Quantity & Time
Naive	1	○	○	○	○	●	○	●	○	○	○	○	S	+	+	–
SMA	2,3	○	○	○	○	●	○	●	○	○	○	○	S	+	+	–
EWMA	4	○	○	○	○	●	○	●	○	○	○	○	S	+	+	–
SES	1,5,8	●	○	○	○	●	○	●	○	○	○	○	S & M	+	+	–
Holt	6,7,8	●	●	○	○	●	○	●	○	●	●	○	S & M	+	+	– / +
Holt-Winters	9,10,11	●	●	○	●	●	○	●	○	●	●	●	S & M	+	+	– / +
ARIMA	1,10,13,14,15	●	●	●	●	●	●	●	○	●	●	○	S & M	– / +	– / +	+
SARIMA	1,12	●	●	○	○	○	●	●	○	●	●	●	M & L	–	– / +	+
ARIMAX	12	●	●	○	○	○	●	●	●	●	●	○	M	–	– / +	+
DLM	14,16,17	●	●	○	○	●	○	●	●	●	●	○	S & M	– / +	– / +	– / +
LASSO	18	●	●	●	○	●	●	○	●	●	○	○	L	–	–	– / +
Random Forest	19,20	●	●	●	○	●	●	○	●	●	○	○	M & L	–	–	+
SVM	21,22	●	●	●	○	●	●	○	●	●	○	○	S	–	–	– / +
SVR	23,24	●	●	○	○	○	●	○	●	●	○	○	S	–	–	– / +
ANN	25,26,27,28	●	●	●	○	●	●	○	●	●	○	○	M & L	–	–	+
ANFIS	29,30,31	●	●	○	○	●	●	○	●	●	○	○	M & L	–	–	+
LSTM	32,33,34,35	●	●	●	●	●	●	●	●	●	●	●	M & L	–	–	+
This Research		●	●	●	●	○	●	●	●	●	●	○	S & M	–	– & +	+

1. Hyndman and Athanasopoulos, 2018	8. Krapp et al., 2013	15. Lee et al., 2021	22. Rostami-Tabar et al., 2013	29. Marx-Gomez et al., 2002
2. Johnston et al., 1999	9. Winters, 1960	16. Ma and Kim, 2016	23. Benti et al., 2023	30. Jin et al., 2013
3. Song and Li, 2008	10. Matsumoto and Komatsu, 2015	17. Geda and Kwong, 2018	24. Yuan et al., 2022	31. Kumar et al., 2014
4. Syntetos and Boylan, 2005	11. Chatfield and Yar, 1988	18. Cui et al., 2020	25. Wang et al., 2019	32. Hochreiter and Schmidhuber, 1997
5. Gardner, 1985	12. Box et al., 2015	19. Suthaharan, 2016	26. Elsaraiti and Merabet, 2022	33. Groenendijk et al., 2020
6. Holt, 1957	13. Turki et al., 2024	20. Sareminia and Amini, 2023	27. Dougherty, 1995	34. Ismail et al., 2018
7. Kumar and Yamaoka, 2007	14. Clottey et al., 2012	21. Zeng and Qiao, 2013	28. Temur and Bolat, 2017	35. Chandriah and Naraganahalli, 2021

Figure 3.4: Matrix of Forecasting Models in Literature Research (Authors' own creation)

4

System Analysis

This chapter aims to analyze ASML's CLSC and RSC processes, focusing on the complete process, part identification and notification systems, and the return flow of defective parts. By examining these components, it identifies key challenges, such as the unpredictability of return volumes and the integration of complex data elements. This chapter will finally give answer to the third sub question of this research:

3. What are the key challenges in managing the return flow of defective parts within ASML's reverse supply chain?

4.1. CLSC & RSC Process Identification

The CLSC at ASML integrates both forward and reverse supply chain processes to optimize the lifecycle management of parts. This integration is achieved by employing methodologies similar to those discussed in the literature review Section 2.1, particularly drawing on the work of Guide and Van Wassenhove (2001) and Thierry et al. (1995). The comprehensive structure of ASML's CLSC is analyzed and depicted using these methods, as illustrated in Figure 4.1.

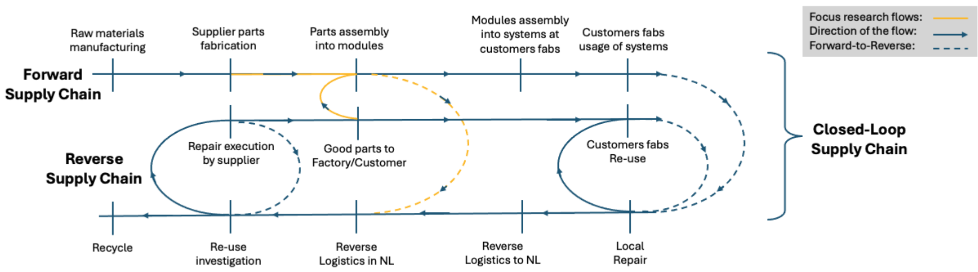


Figure 4.1: General CLSC of ASML (Authors' own creation)

ASML's supply chain starts with external suppliers manufacturing raw materials, followed by the fabrication of parts by both internal and external suppliers. In the forward supply chain of ASML, these parts are assembled into modules, which are then integrated into complete systems at semiconductor fabrication plants, known as fabs. Effective planning ensures the timely availability of spare parts, minimizes production downtime, and fulfills customer orders. Each part is identified by a unique numerical code and assembled into modules for lithography systems.

ASML engages in recovery operations mainly for economical and circularity reasons, as refurbishing parts is more cost-effective than purchasing new ones while maintaining high quality. The RSC consists of two main inflows: return flow from the factory and return flow from customer fabs. Returned parts from customers, known as service parts, may be repaired locally if facilities exist; otherwise, they are sent to the Netherlands for further processing if it is economically and sustainably viable. This process includes repair execution, returning good parts to factories or customers, RL within the Netherlands, repair and reuse investigations, recycling, and local repairs. This closed-loop approach supports ASML's sustainability and operational excellence goals, ensuring economic objectives are met while minimizing waste and optimizing resource reuse.

As indicated in the scope section, this research focuses on parts that become defective within the factory. The primary issue concerns the inflow of new and repaired parts from suppliers and their subsequent assembly. The inflow of parts from the reverse inventory's good stock of repaired parts indirectly influences the scope of this research. The core aspect of the return inflow into the reverse inventory is the outflow of these defective parts from the factory; these flows are indicated by the orange lines in Figure 4.1. Repairable parts are returned to suppliers for repairs, while unrepairable parts undergo a reuse investigation to determine the appropriate action, such as dismantling, harvesting, or recycling. This dynamic necessitates identifying the variables essential for developing a forecasting model as part of ASML's RSC.

Figure 4.2 provides a simplified forward flow of ASML's supply chain, highlighting the factory parts and service parts for customer facilities. The red "Defect" cross indicates where defects occur in the supply chain and the point at which forecasts must predict these defects. This figure clarifies that ASML also acts as an internal supplier of parts. Initially, these raw materials are supplied by external suppliers and then developed into unique ASML parts using several external components. These parts are distributed to factory warehouses for use or to service warehouses. The indicated gray shapes are outside the scope of this research. Additionally, the forward flow to the factory is indirectly connected to the return flow of defective machinery parts. This connection helps in understanding the relationship between these flows and gathering data variables.

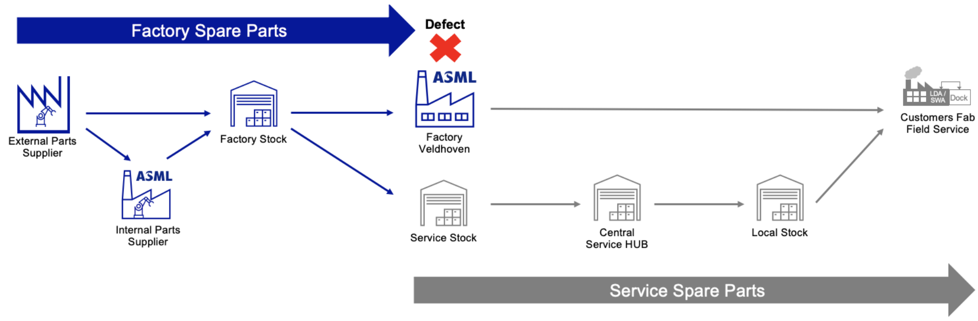


Figure 4.2: Forward Flow of ASML Supply Chain (Authors' own creation)

Figure 4.3 shows a simplified reverse flow from the factory in ASML's supply chain, primarily investigated to understand the consequences and identify the problem statement. The red "Defect" cross marks the points in the supply chain where defects occur and where forecasts must predict these defects. This indicates that the defect point is where the forward flow transitions into the reverse flow. The reverse flow of defective machinery parts from the factory leads to overcapacity in the reverse inventory, and repairable parts are sometimes recycled due to a lack of demand, leading to value loss. Repairable defective parts are stocked in the 5L warehouse and only sent to the supplier for repairs if there is a demand trigger from the factory or customer facilities. Unrepairable parts, or those without current demand, are sent for evaluation for potential reuse, provided that warehouse capacity allows. When there is demand for repairable parts, they can eventually be sent to the supplier for repairs. Unrepairable parts follow a distinct process, depicted in grey, leading to decisions regarding dismantling, harvesting for usable components, or recycling. The dismantling decision is entirely outside the scope of this research.

The repair process and inventory control, although outside the direct scope of this research, are intricately connected to it, and their interdependencies will be reflected in the recommendations. The demand trigger in the reverse flow is crucial for conducting a classification analysis to identify high-value parts for this case study. Furthermore, understanding lead times for new parts in the forward flow and repairable defective parts in the reverse flow is valuable for determining the appropriate forecasting horizon for each part. This approach align with the problem statement's recommendation to block new buys and create dummy orders for repairable defective parts in the system.

4.2. Part Identification and Notification Systems

This subsection delves into ASML's part identification and notification system, which is essential for subsequent analyses. Each component at ASML is assigned a unique 12-digit Numerical Code (12NC) along with a specific equipment number that facilitates individual tracking. When modifications or defects

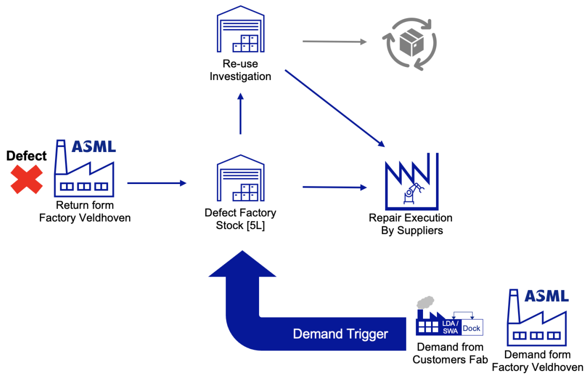


Figure 4.3: Reverse Flow of ASML Supply Chain (Authors' own creation)

occur, a MN number is generated, which plays a critical role in the precise tracking of each part's lifecycle. Grasping this system is crucial for the logistical and supply chain operations at ASML and lays the groundwork for further analytical investigations. Figure 4.4 illustrates the straightforward relationship between these three identification and notification elements, with arrows to the right indicating multiple 12NC lines comprising individual parts and various material notifications.

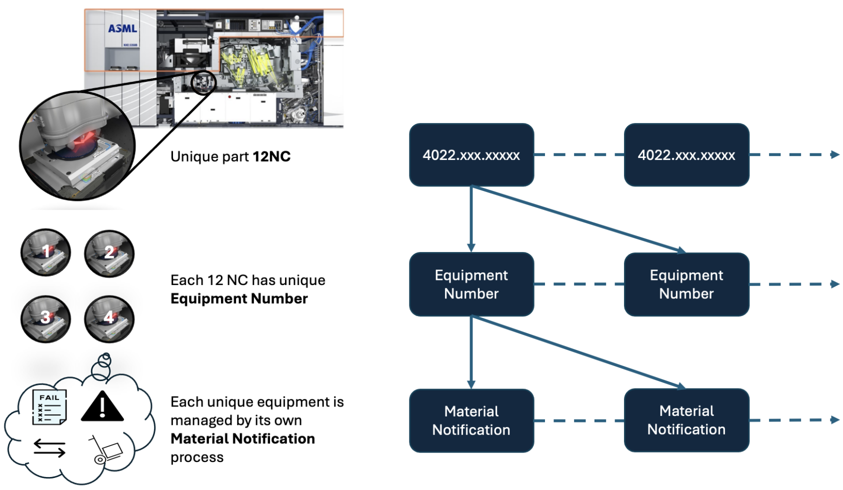


Figure 4.4: 12NC, Equipment Number and MN (Authors own creation)

4.2.1. 12-digit Numerical Code (12NC)

At ASML, the management of part inventories and the tracking of part modifications are facilitated by the 12NC system. This standardized system is crucial for identifying parts across various databases, including System Applications and Products in Data Processing (SAP) a widely-used enterprise resource planning software and Teamcenter a product lifecycle management database system enhancing logistical efficiency and operational accuracy.

ASML's parts, which often exhibit non-trivial complexity and significant value, are sourced globally from multiple suppliers. These parts, whether as larger modules or individual spare components, are managed through a robust identification system using the 12NC, essential for efficient lifecycle management of each component.

Each ASML machine component is designated by a unique 12NC, an important legacy from ASML's Philips heritage. The 12NC is integral to precise component tracking and effective inventory management. Below is a detailed breakdown of the 12NC structure:

- **First 4 Digits:** Specify the site of part usage. For instance, '4022' denotes parts used in ASML factories, while 'SERV' indicates parts designated for customer service sites. This research focuses on '4022' due to its relevance to factory operations.

- **Next 3 Digits:** Relate to the department or project, e.g., '657' for Mechanical EUV projects.
- **Following 4 Digits:** Are uniquely assigned to each part, ensuring specific identification.
- **Last Digit (Modification Stage):** Tracks the modification history of the part, starting at '1' for new designs and increasing with each modification.

These 12NC identifiers are generated through the Material Reservation Tool (MRT) and integrated across various ASML systems for enhanced tracking and management.

Similarly, the 11-digit Numerical Code (11NC) comprises the first 11 digits of the 12NC as the base part number, enhanced by a 12th versioning digit. This versioning digit, which ranges from 1 to 9, indicates a part's direct lineage to its predecessors or successors, showcasing the evolution of the part through modifications. When the number of modifications exceeds nine, a new 11NC is initiated, signifying a shift in the base part number. In contrast, each 12NC number is completely unique, acting as a standalone identifier for parts that typically do not undergo frequent modifications. This ensures that each part remains uniquely identifiable across all ASML systems.

This structured approach to part numbering via the 11NC and 12NC systems is crucial for maintaining an orderly and effective inventory and manufacturing process management, ensuring each part's history and modifications are clearly tracked and managed.

4.2.2. Equipment Numbers

At ASML, each part is not only identified by a 12NC but also assigned an equipment number. This equipment number acts as a unique serial identifier for each individual part, linking it to its specific 12NC. It enables a detailed tracking of where each component is used, how it interacts with different assemblies, and its function within various systems. This system ensures precise tracking and management of parts throughout their lifecycle.

The importance of the equipment number extends beyond simple identification. It is crucial for maintaining records of calibration, certification, and compliance with quality standards, all of which are essential elements in the precision-demanding field of semiconductor manufacturing. Furthermore, equipment numbers enable maintenance teams and engineers to quickly locate and identify the correct parts for repairs or upgrades, thus minimizing downtime and enhancing operational efficiency.

Each part's lifecycle is recorded under its equipment number in ASML's internal databases. Actions such as "Equipment created", "Outbound delivery", and "Functional location changed" are constantly tracked. This lifecycle data is not only vital for quality assurance but also helps factory Material Quality (MQ) engineers monitor the parts for any unusual behavior or potential failures. By analyzing years of lifecycle data, engineers can identify patterns or anomalies that might indicate a risk of failure, allowing for proactive measures to enhance reliability and operational safety.

4.2.3. Material Notification MN Process

The MN creation process at ASML is a system designed to manage updates, replacements, or checks on inventory involving specific parts, identified by their unique 12NC and associated equipment numbers. This process is essential for maintaining operational integrity and ensuring the quality of part modifications and replacements. The MN process unfolds through several carefully planned steps:

- **Notification Creation:** Whenever a part is newly introduced or requires modifications due to upgrades or defects, a MN is generated. This document is detailed, specifying whether the part is being replaced, modified, or introduced for the first time. It is crucial for maintaining accurate records of parts and their conditions.
- **Notification Approval:** After the creation of the MN, it must be approved by key stakeholders, such as supply chain managers, production engineers, and quality assurance teams. This approval is critical to ensure that all changes adhere to ASML's stringent standards and specifications, ensuring each component conforms to the system's overall quality framework.
- **Implementation in Systems:** Once approvals are obtained, the specifics documented in the MN are integrated into ASML's inventory and management systems. This step is fundamental to keep

all inventory records up to date and ensure that the system databases accurately reflect the latest status of parts and components, essential for the seamless operation of the company.

This research highlights the importance of generating a MN when a defect is detected in a part. The creation date of the MN is recorded and linked to the defect's occurrence, which is crucial for accurate forecasting and data analysis. When a defect is detected, an MN is created to capture essential details, such as the plant code, equipment number, and the specific system affected. The process of MN creation and the subsequent steps are illustrated in Figure 4.5, and the key steps are shortly described:

- **Root Cause Analysis (RCA):** This process involves a detailed investigation to determine the underlying cause of the defect *(Validate failure analysis)*. Strategies are then developed to prevent future occurrences. The MN remains active until all corrective measures are implemented, ensuring a comprehensive resolution of the issue.
- **Material Recovery and Closure:** The Material Review Board (MRB) evaluates the defect and determines the most suitable corrective action: *in-house repair, return to the supplier, or scrap the material*. This decision follows established recovery indicators. Additionally, the MRB checks the accuracy and completeness of the MN data. The MN remains open until the recovery actions and RCA are fully completed.
- **Lifecycle Updates and Vendor Feedback:** Information about the lifecycle of defective parts, including any repairs or replacements, is precisely recorded in the MN. Feedback from vendors regarding the condition of returned parts offers essential insights for potential improvements and helps identify recurring issues, thereby contributing to continuous quality assurance. If a part is deemed non-repairable upon receipt, it undergoes another inspection, and a new MN is created to document this status.

This MN approach ensures that each stage, from defect identification to resolution, is carefully monitored and documented. This not only upholds high standards of quality and reliability but also minimizes operational disruptions within ASML's complex supply chain.

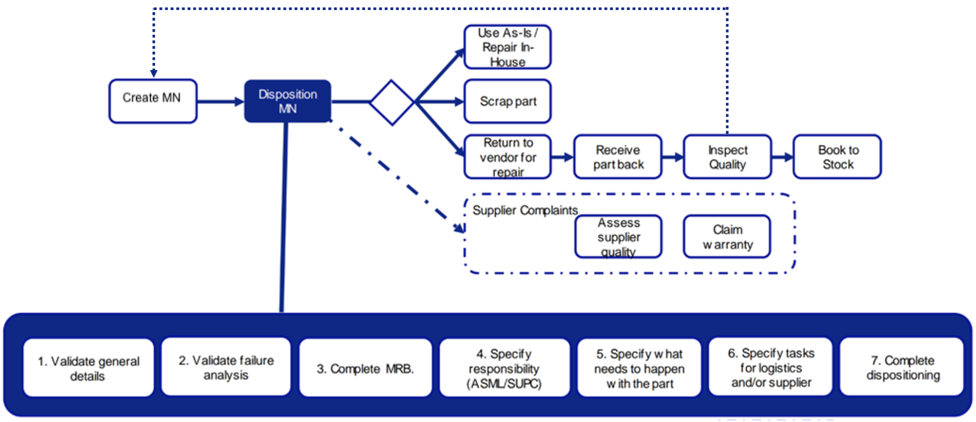


Figure 4.5: Material Notification Creation (ASML's Process Modified by Author)

4.3. Swimlane Process Analysis

The Swimlane diagram in Figure 4.6 visually represents ASML's process for managing defective parts, primarily within the RSC, but also as part of the complete CLSC. The diagram is divided into five main columns: *Demand & Planning, Supplier/Vendor, Factory Veldhoven ASML, MRB, and Output*. Each column corresponds to a distinct department or the output phase in the process flow, illustrating their interactions and dependencies. For a more detailed view, an enlarged version of the Swimlane diagram is available in Appendix B.

- **Demand & Planning**: The process begins with the *Demand & Planning* departments, where a demand forecast triggers the initiation of a "12NC Factory Part Process". During this phase, it is determined whether a repairable part is available in stock. If such a part is available, it is sent back to

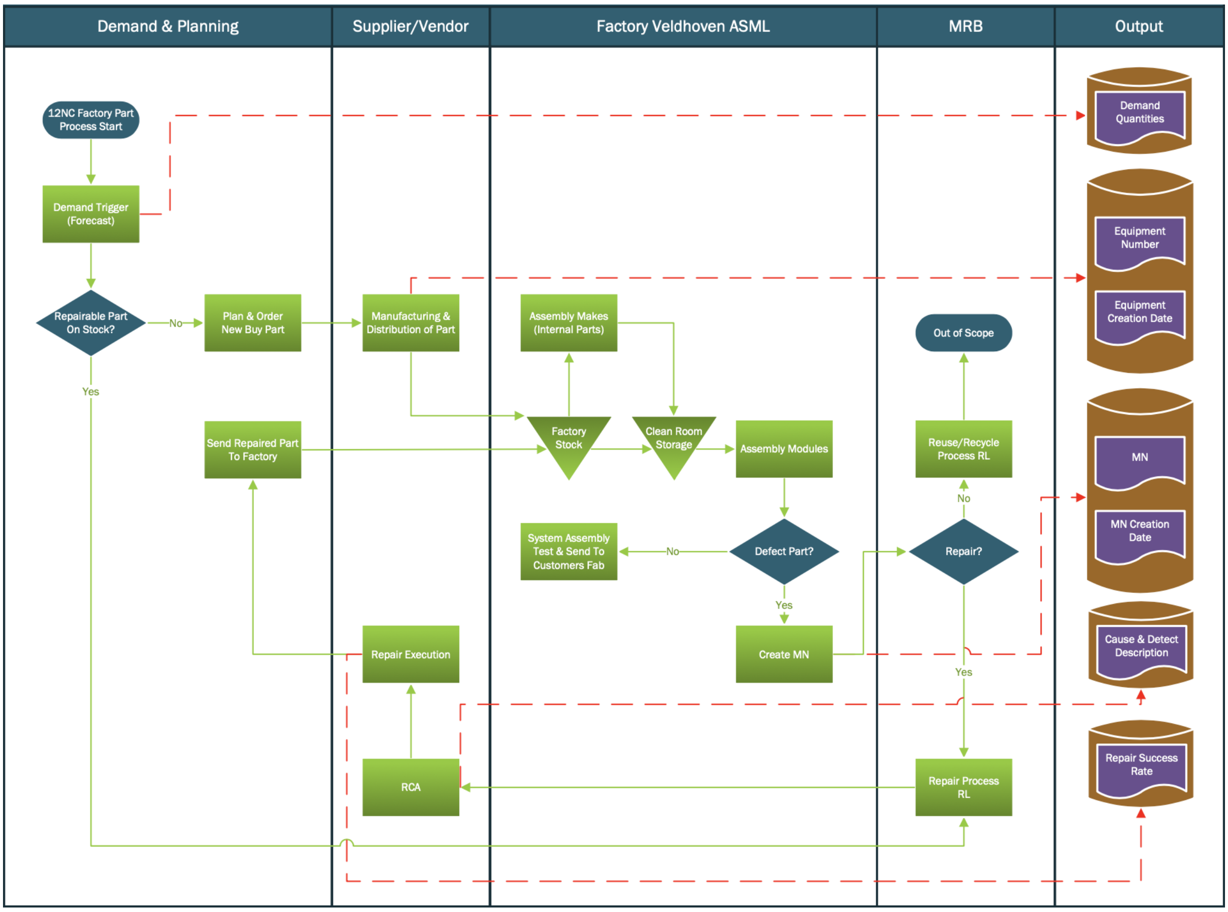


Figure 4.6: Swimlane Process of Defective Parts (Authors' own creation)

the factory for further processing. If no repairable part is available, a new part is planned and ordered to meet the demand.

- **Supplier/Vendor**: The suppliers, mainly vendors, are responsible for the manufacturing and distribution of parts. These parts are then either stored as factory stock or moved to the next phase of assembly. This stage ensures that the necessary components are available for the production process at the *Factory Veldhoven ASML*.
- **Factory Veldhoven ASML**: At the *Factory Veldhoven ASML*, the operations involve several critical steps. Parts are initially stored in clean room storage before being assembled into modules. The ASML operations can be categorized in module assembly (ASSY) and systems assembly (FASY). These modules undergo system assembly and testing, after which they are sent to customer fabs. If a defect is detected during any phase of this process, a MN is created. The defective part is then subjected to repair execution and RCA.
- **MRB**: The MRB assesses whether a defective part can be repaired. If a part is deemed repairable, it undergoes the repair process and is either reused or recycled. Non-repairable parts are out of scope of this analysis, ensuring that only viable components are reintegrated into the supply chain. The repairable parts process is connected to show the RCA connection.
- **Output**: The Output phase captures critical data points essential for effective forecasting and SCM. These data points include demand quantities, equipment numbers and creation dates, MN's and their creation dates, cause and defect descriptions, and repair success rates. By analyzing these data points, ASML can improve forecasting accuracy, optimize inventory management, and enhance overall supply chain efficiency. This structured approach ensures that the RSC process is well-managed, reducing downtime and maintaining high operational standards.

Key Data Outputs

The Swimlane diagram highlights several essential data outputs, as explained below.

- **Demand Quantities**: Understanding the volume of spare parts demanded is valuable for determining the classification of parts and identifying which defective parts are significant for the case study.
- **Equipment Number and Creation Date**: Facilitates tracking and lifecycle analysis of parts, aiding in better asset management.
- **Material Notifications**: Provides detailed records of defects and issues, helping to identify the count and timing of parts becoming defective.
- **Cause and Detect Descriptions**: Assists in identifying common defect patterns and root causes, enabling proactive measures.
- **Repair Success Rate**: Indicates the efficiency and effectiveness of repair processes, offering valuable insights for recommendations regarding specific defective parts.

By thoroughly analyzing these data points, they can serve as crucial inputs for categorization in forecasting models within this process.

Defective Causes of Parts in Factory and Quality Issues

Defective parts in the early stages of factory processes can significantly impact ASML's supply chain efficiency and product quality. These quality issues stem from various factors, including substandard raw materials, inconsistent manufacturing processes, and inadequate quality control measures from suppliers. Additionally, transportation mishaps such as handling damage, environmental exposure, and packaging failures further contribute to defects. Human errors during assembly and equipment malfunctions also play a critical role. These defects lead to production delays, increased repair costs, and potential customer dissatisfaction.

Upon being stored in the "Factory Stock" as shown in the Swimlane diagram, defective parts undergo investigation. During the sub-assembly process, where parts are combined to create new ASML components referred to as "Makes" and during the cleaning of the parts in the "Clean Room Storage" inspections determine their usability. This process helps identify whether defects originated from the supplier or occurred during transportation. If the parts are deemed good stock, they proceed to assembly. However, defects can still arise during processing in the "Assembly Modules" or due to design faults.

4.4. Current vs Forecast-Driven Approach Comparison

Currently, ASML uses a reactive method for ordering parts in the factory, primarily issuing new buy orders when defective parts are unavailable for repair. This reactive approach results in inefficiencies, overstocking, and extended lead times. Table 4.1 illustrates the differences between the current reactive approach and a forecast-driven approach to operational factory parts management, highlighting the potential improvements in efficiency, inventory control, and lead time reduction that a forecast-driven strategy can offer.

4.5. Conclusion

This section synthesizes the research outcomes from the system analysis as addressed in this chapter, specifically focusing on the third sub-question of this research:

Research Question 3

What are the key challenges in managing the return flow of defective parts within ASML's reverse supply chain?

The focus is to analyze the complexities in managing the return flow of defective parts within ASML's RSC. As a subset of the broader CLSC, where parts and modules flow cyclically, ASML's RSC specifically addresses the causes, return flow, and inventory of defective parts originating from the factory, presenting unique challenges.

A major challenge is the unpredictability of return volumes. This unpredictability starts from when a part is identified as defective, documented through a 'Material Notification' (MN), which is crucial for initiating the

Aspect	Current Approach	Forecast-Driven Approach
Demand Management & Order Planning	• Reactive with immediate new buy orders • Inefficient planning, unreliable forecasts, leading to oversupply	• Proactive, anticipating defective parts returns • Efficient planning, reliable forecasting, reducing new buy orders
Lead Times & Operational Efficiency	• Long repair lead times, significant delays • Inefficient operations, high costs, overstocking, underutilized repaired parts	• Shorter lead times with proactive planning • Cost-effective, sustainable operations • Improved inventory management, faster repair cycles
Inventory & Cost Management	• High dependency on new buy orders, leading to overstocking and increased costs	• Reduced dependency on new buy orders, focusing on the repair loop • Cost savings, better inventory management

Table 4.1: Comparative Analysis of Current vs. Forecast-Driven Approaches at ASML

return process. Without predictive insight into future return volumes, specifically the quantity of parts that will become defective, inventory management within the RSC becomes significantly complicated. This unpredictability makes maintaining optimal stock levels and planning for new purchases difficult, leading to potential overstocking or stockouts.

Additionally, managing the return flow is complicated by the need to effectively integrate key data elements such as 12NC, equipment numbers, and MNs. These identifiers are crucial for tracking defective parts throughout their lifecycle. However, the complexity and vast number of different parts across various modules and systems present significant challenges. The decision-making process for handling defective parts requires precise demand triggers and inventory controls. Understanding lead times for new parts and repairable defective parts is crucial for determining the appropriate forecasting horizon. Aligning inventory levels with actual demand is essential to minimize unnecessary new purchases.

To address these challenges, enhancing predictive insights through the analysis of historical return patterns and developing reliable forecasting models tailored to the RSC is essential. Integrating key data outputs, such as demand quantities, equipment creation dates, cause and defect descriptions, and repair success rates, can improve tracking and management capabilities. This integration will enable ASML to implement effective forecasting models, enhance inventory management, and ensure that the RSC operates more efficiently.

5
Data Analysis

This chapter aims to provide a comprehensive analysis of the data available at ASML for forecasting defects in machinery parts, specifically focusing on the Veldhoven factory. The primary goal is to answer the fourth sub-research question of this thesis:

4. What data is available at ASML for predicting the return flow volume of defective parts, and how can this data be processed with a focus on the crucial variables?

To achieve this, the chapter explores the data systems and tools employed at ASML, detailing the processes of data extraction, filtering, and merging. It also discusses the methods used for data preparation and understanding. Finally, the data will be transformed to ensure the datasets are suitable for accurate time series forecasting. The overall process of data analysis, from initial data extraction to final data transformation, is illustrated in Figure 5.1.

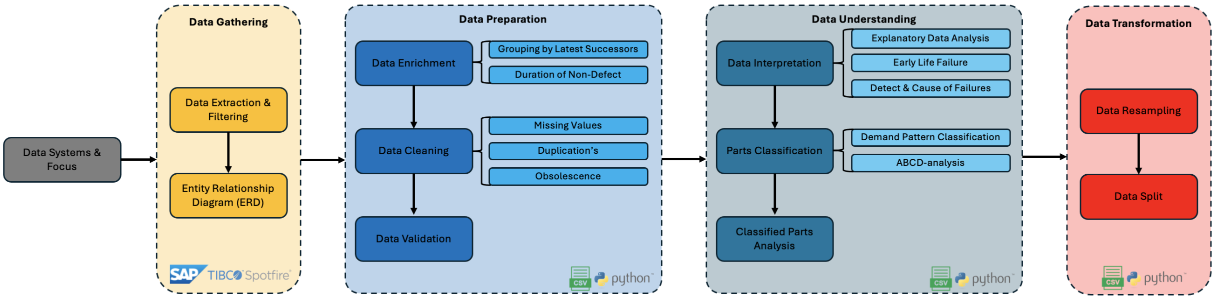


Figure 5.1: Data Analysis Process Overview (Authors' own creation)

5.1. Data Systems & Focus

The primary objective of this research is to examine forecasting models that accurately predict the return flow volume of defective machinery parts within the Veldhoven factory. To achieve this, the research aims to forecast when specific 12NC parts will fail by using historical MN data for different equipment numbers. Each MN creation date, obtained from SAP ECC and processed in Spotfire, serves as a timestamp indicating the occurrence of a defect. By analyzing and aggregating these timestamps, a forecasting model could predict the total number of expected defective instances for each unique 12NC part over a specified period.

ASML employs the SAP ECC system to gather and process MN data, which is crucial for RSC operations. The primary SAP dataset for this research is ZLifecycle, a transaction system accessible via specific equipment numbers. ZLifecycle documents the lifespan of ASML parts, excluding simpler parts like bolts, and includes attributes such as 12NC, plant location, part description, and events detailing what happened with the specific equipment. Each MN linked to an equipment number includes a creation timestamp. Further details on these lifecycle data attributes will be explored in the following sections.

To analyze and visualize data, ASML's RSC Ops team uses TIBCO Spotfire, an AI-enhanced analytics platform. Spotfire integrates data from SAP through the SAP HANA connector, which combines static data with real-time data from various sources. These connectors, also known as "infolinks" enable seamless data integration. Spotfire provides dynamic dashboards and interactive visualizations for comprehensive big data analytics. Data from SAP ECC ZLifecycle is continuously streamed into Spotfire, where it is collected and reorganized. For this research, the primary Spotfire dataset used is "CSAP Material Notification" which gathers ZLifecycle data from SAP. Further analysis will involve exporting this data to Comma-Separated Value (CSV) files for additional cleaning, processing, and modeling. The process of extracting data from SAP through a data warehouse into a Spotfire dashboard will be detailed further.

For data cleaning, processing, and developing the forecasting model, this research utilizes an ASML laptop equipped with an Intel Core i7 processor. Python version 3.9 is used as the programming language on the Jupyter Notebook platform, employing packages such as `NumPy`, `Pandas`, `scikit-learn`, `Seaborn`, and `Matplotlib`, all installed in a virtual environment. The choice of Python is due to the researcher's familiarity with the language and its efficiency in handling large datasets, which is essential for this research. However, other programming languages could also be suitable for this task. The data exported from Spotfire is stored as CSV files and managed using Python during the preparation and modeling phases. The data flow connections of these different data systems are illustrated in Figure 5.2.

Figure 5.2: Connection Between Data Systems (Authors' own creation)

5.2. Data Gathering

This section outlines the initial steps undertaken to gather and pre-process the data for analysis. The process encompasses the extraction and filtering of relevant data from ASML's systems, followed by the creation of a merging diagram to illustrate the connections between datasets. These pre-processing steps ensure that the data is systematically organized and refined, thereby facilitating subsequent preparation and understanding of the data.

5.2.1. Extraction & Filtering

Data extraction from Spotfire starts with the importation of the "CSAP Material Notification" information link. This process involves setting filters on demand data to narrow down the relevant data points for this research. The filtering options are considered as follows:

- **Date of Notification:** This corresponds to the creation date of the MN. Preference is given to entries no older than five years to ensure relevance and accuracy. Dates in this report are presented in the European format: DD/MM/YYYY. The data range is limited to records from the past five years, as data from this period is deemed the most qualitative and reliable. Older data becomes less relevant for forecasting. The reference point is the most recent export date, 31/05/2024, which is chosen because data is updated on the last day of each month. This snapshot was made on 06/06/2024. Thus, the data covers the period from 31/05/2019 to 31/05/2024.
- **Plant for Material:** The research focuses on the Veldhoven factory, which represents the early lifecycle stages of parts. The plant code for this location is "NL01" which is used as a filtering parameter to scope the data accordingly.
- **Material Notification Quality Sufficient:** Given that MN creation is a manual process and subject to human error, each MN is double-checked by Material Quality (MQ) engineers. A parameter is set to select MN's where the quality is marked "Y" for sufficient, ensuring only high-quality notifications are included in the dataset.
- **Notification Type:** This parameter specifies the MN type. By setting the constraint to "ZQ" the dataset is limited to MN's that report failures, aligning with the research focus on defective parts.
- **Detect Code:** This parameter identifies the detection code associated with an MN. Setting the constraint to "ZQ-MESDT" ensures inclusion of only those detection codes relevant to the Veldhoven plant. These codes provide insights into the primary causes of MN creation, such as "F: Failed during processing."

After implementing these filters, the dataset comprises 83,211 rows, representing the total of MN's. Additionally, new columns have been integrated into the Spotfire Data Canvas environment, followed by specific row filtering to eliminate irrelevant MN's:

- **Factory Clean 12NC:** A new column is calculated using an expression that assigns "Y" only to 12NC codes that strictly begin with "4022.xxx.xxxxx" and contain no additional characters. This filter effectively excludes special cases like SERV, USP, UPP, and FSD from the dataset, as they are beyond the scope of this research. Retaining only the rows marked "Y" cleans the dataset, reducing it to 78,701 rows.
- **Equipment Number CSAP Clean:** A new column is calculated using an expression that assigns "Y" only to equipment numbers from SAP that are strictly between 7 and 14 characters in length. This filter effectively excludes all parts that are missing an equipment number or are incorrectly registered in SAP. Retaining only the rows marked "Y" cleans the dataset, reducing it to 68,947 rows.

The data is now initially filtered from a scope perspective. The "CSAP Material Notification" information originally consisted of 164 columns; after adding the new calculated columns, it now comprises 166 columns. Each column represents an attribute, which is a data field describing a characteristic or feature of a data object associated with a specific MN.

5.2.2. Entity Relationship Diagram

Collecting data and understanding the interconnections within ASML's complex systems posed significant challenges. This process required consulting with various experts to gather insights and make informed decisions based on the literature review. As illustrated in Figure 4.6, the Swimlane supply chain process outlines the potential data outputs, which are further discussed in Section 4.5. This data output can be further enhanced by incorporating additional relevant attributes from various tables in SAP and Spotfire, applying these in potential forecasting models. The most critical columns from the "CSAP Material Notifications" were selected to reduce dataset complexity, with the option to add more columns in the future if necessary. The following datasets are extracted and filtered to expand the dataset for this research: "SAP IH10 Equipment Selection", "Material Masterdata On Plant Level", "ReUse Demand", and "RSR (Repair Success Rate)".

The Entity Relationship Diagram (ERD) in Figure 5.3 illustrates the detailed structure and connections between the datasets within ASML's systems to merge the data tables. These tables are interconnected through key fields such as material number (12NC) and equipment number, enabling comprehensive data merging essential for accurate forecasting models. The "Notification Number" refers to Material Notifications (MN). The cardinality and relationships between "CSAP Material Notifications" and these tables are explained below.

- **SAP IH10 Equipment Selection:** This relationship is many-to-one. Each equipment number in the "SAP IH10 Equipment Selection" table can be associated with multiple MN's in the "CSAP Material Notifications" table. This indicates that one piece of equipment can generate numerous MN's over time, capturing a detailed history of defect activities. This dataset is named "IH10" and is indirectly extracted, first from SAP ECC to a CSV file and then imported into Spotfire.
- **Material Masterdata On Plant Level:** This relationship is many-to-one. Each MN in the "CSAP Material Notifications" table is linked to a single material number (12NC) entry in the "Material Masterdata On Plant Level" table. This means that while many MN's can refer to the same 12NC, each 12NC entry consolidates all related MN's, providing a centralized view of material specific attributes. This dataset is extracted directly from Spotfire with an infolink.
- **ReUse Demand:** This relationship is one-to-one. Each 12NC in the "ReUse Demand" table is directly linked to a corresponding MN in the "CSAP Material Notifications" table. This ensures that the reuse demand data is specific and directly correlates to individual MN's, providing precise demand insights. This dataset is extracted directly from Spotfire with an infolink.
- **RSR (Repair Success Rate):** This relationship is one-to-one. Each MN in the "CSAP Material Notifications" table is connected to a unique entry in the "RSR" table, which tracks the success rates of repairs for each 12NC. This linkage allows for detailed analysis of repair outcomes for each notified defect, aiding in the assessment of repair efficacy. This dataset is extracted indirectly from a CSV file into Spotfire.

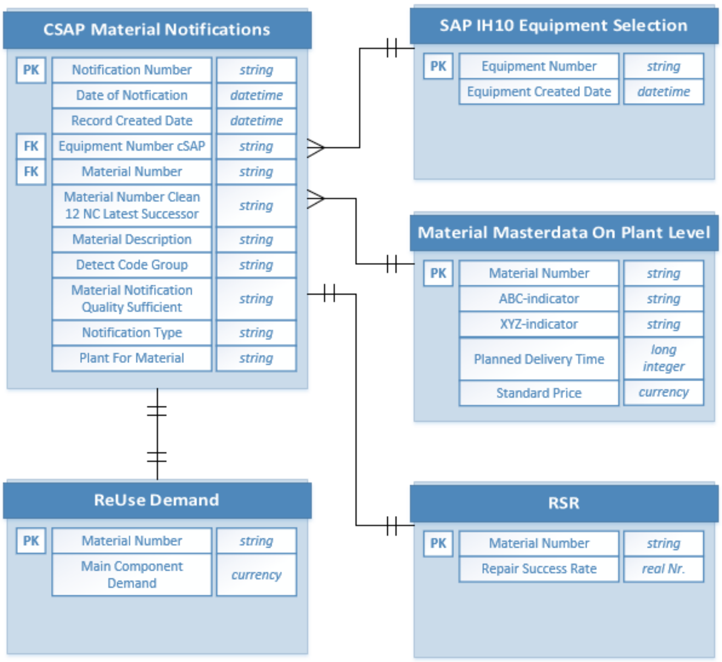


Figure 5.3: Entity Relationship Diagram of Data Merging (Authors' own creation)

This process results in a consolidated dataset comprising 19 columns, constructed from the integration of five distinct datasets. The number of rows remains unchanged, as the merged dataset continues to be based on the MN level.

5.3. Data Preparation

The original dataset, after extraction, filtering, and merging, was thoroughly investigated. The summary of all column attributes, including non-null counts, null counts, and unique counts, is shown in Table 5.1.

5.3.1. Data Enrichment

To enhance the reliability of forecasting models, the dataset was enriched by implementing two key strategies as detailed below.

Grouping by Latest Successors

During the data preparation phase, the initial focus was on the original 12NC parts identified by their `Material Number`. However, frequent upgrades within a two-year period often resulted in predecessors and successors sharing many characteristics. One main similarity is that when one 12NC is unavailable, a different version with a small upgrading adjustment can be used for the same operational task, leading to a similar demand quantity for these parts. To address this relationship, the dataset was enriched by grouping parts by their latest successors, represented by the `Material Number Clean 12NC Latest Successor`. This approach aggregates similar parts that have become defective over different years, maintaining a high degree of relatedness.

Consultations with ASML's experts confirmed the substantial relationship between these parts, justifying the decision to group them by their latest successor. The original dataset contained 12,201 unique `Material Number` values, while the enriched dataset contains 9,832 unique values. This reduction in unique 12NC's effectively increases the defect counts per group, providing a more robust dataset for analysis. Consequently, this enrichment enhances the reliability of the forecasting models by allowing for more historical data points per part group, resulting in improved predictive accuracy.

Duration of Non-Defect

To further enrich the dataset to analyse the defective parts, an additional feature attribute, `Duration of Non-Defect`, was created. This feature represents the time interval between the `Equipment Created Date`

#	Column Name	Non-null	Null	Unique
1	`Notification Number`	68,947	0	68,947
2	`Date of Notification`	68,947	0	1,800
3	`Record Created Date`	68,947	0	1,799
4	`Equipment Number CSAP`	68,947	0	61,317
5	`Material Number`	68,947	0	12,201
6	`Material Number Clean 12NC Latest Successor`	68,856	91	9,832
7	`Material Description`	68,947	0	9,744
8	`Detect Code Description`	68,947	0	9
9	`Source Of Cause Code Description`	67,940	1,007	15
10	`Material Notification Quality Sufficient`	68,947	0	1
11	`Notification Type`	68,947	0	1
12	`Plant For Material`	68,947	0	1
13	`Equipment Created Date`	68,933	14	3,757
14	`ABC-indicator`	63,578	5,369	3
15	`XYZ-indicator`	49,077	19,870	3
16	`Planned Delivery Time`	68,937	10	104
17	`Standard Price`	68,604	343	11,014
18	`LatestSuccessorDemand`	54,530	14,417	738
19	`RSR`	48,067	20,880	219

Table 5.1: Summary of Data Columns with Non-Null Counts, Null Counts, and Unique Counts

and the `Date of Notification` for each defect. The creation of this feature involved several key steps:

1. **Date Conversion**: Converted the `Date of Notification` and `Equipment Created Date` from string format to datetime objects, enabling accurate calculations.
2. **Calculation of Duration**: Calculated the duration of the defect by subtracting the `Equipment Created Date` from the `Date of Notification`.
3. **Identification of Negative Durations**: Identified instances with negative durations, indicating potential data entry errors or anomalies.
4. **Verification of Data**: Assessed the count of negative duration values, finding 276 instances, and reviewed a sample to verify the correctness of the `Duration of Non-Defect` feature.

The `Duration of Non-Defect` feature adds significant value by providing insights into the time elapsed between the creation of equipment and the notification of defects. This feature can be utilized for predictive modeling, quality control, and data integrity checks, significantly enhancing the overall analysis.

5.3.2. Data Cleaning

The data cleaning process involves several crucial steps to ensure data integrity and validity. Starting with 68,947 rows, the dataset was filtered down to 44,718 rows by addressing missing values, unrealistic data points, and critical field completeness.

Missing Values

The data cleaning process for this dataset involved several steps designed to ensure the integrity and validity of the data. Each step was carefully chosen to address specific data quality issues. The dataset cleaning missing values resulted in a total of 20,728 rows being dropped. Table 5.2 provides a detailed explanation of each step and the rationale behind the choices made.

The data cleaning process was essential to create a valid and reliable dataset for analysis. By addressing missing values, unrealistic data points, and ensuring the completeness of critical fields, the resulting dataset is robust and suitable for subsequent analysis. These steps ensure that the dataset accurately reflects the real-world scenarios it aims to model, providing a solid foundation for further insights and decision-making.

Step	Action	Rows Affected
1	**Drop rows** with missing values in `LatestSuccessorDemand`	14,417
2	**Drop rows** where `LatestSuccessorDemand` > 900,000	5,864
3	**Drop rows** with missing values in `Standard_price`	287
4	**Drop rows** with missing values in `Equipment Created Date`	13
5	**Drop rows** with negative `Duration of Non-Defect`	147
6	**Mark** missing values in `Source Of Cause Code Description` as `Unknown`	636

Table 5.2: Summary of Missing Values Steps

Duplication's
To maintain the accuracy and reliability of the dataset, it is essential to identify and remove duplicate records. Duplicates can arise due to various reasons, such as data entry errors or system glitches, and can negatively affect the analysis by skewing results. In this study, duplicates were identified based on three key columns: `Date of Notification`, `Equipment Number CSAP`, and `Notification Number`. The process was carried out as follows:

1. The dataset initially contained 48,219 rows.
2. Groups with multiple MN's for the same equipment number on the same date were identified.
3. Within each group, all but the first occurrence were marked for removal to eliminate redundant entries.
4. This process resulted in the removal of 804 rows, reducing the dataset to 47,415 rows.

By removing these duplicates, the dataset was refined to more accurately represent the true MN events, thereby enhancing its integrity and the reliability of subsequent analyses.

Obsolescence
To further ensure the dataset's relevance and reliability, it is necessary to identify and remove records of the parts that have not had any defect notifications in the last two years. These parts are considered obsolete, either because their defects have been resolved or they are no longer in use. Removing these parts helps maintain an accurate dataset for analysis and forecasting. The process was carried out as follows:

1. The dataset initially contained 47,415 rows.
2. Convert the dataset's `Date of Notification` field to a datetime format.
3. Define the cut-off date as two years prior to May 31, 2024.
4. Group the dataset by `Material Number Clean 12NC Latest Successor`.
5. Identify groups with at least one defect recorded within the last two years.
6. Mark parts without any defect records in the last two years, resulting in 1,039 parts (12NC).
7. Remove the rows of these parts from the dataset, dropping 2,697 rows, resulting in 44,718 rows after cleaning.

By removing obsolete parts, the dataset was refined to focus on relevant and active defects, thereby enhancing its integrity and the reliability of subsequent analyses.

5.3.3. Data Validation

The dataset's integrity and cleanliness are critical for accurate analysis. Table 5.3 provides a comprehensive overview of the dataset after the cleaning process. Notably, there are no missing values, except for three attributes which are not required for further data analysis. These attributes, however, may be valuable for elaboration or recommendations. Additionally, the dataset's uniqueness has been validated, ensuring that all entries are distinct and reliable.

5.4. Data Interpretation

This section provides an in-depth analysis of the processed data to extract meaningful insights and understand the underlying patterns. It involves examining trends in defect notifications, identifying early life failures, and analyzing the causes of failures.

#	Column Name	Non-null	Null	Unique
1	`Notification Number`	44,718	0	44,718
2	`Date of Notification`	44,718	0	1,776
3	`Record Created Date`	44,718	0	1,773
4	`Equipment Number CSAP`	44,718	0	40,345
5	`Material Number`	44,718	0	5,845
6	`Material Number Clean 12NC Latest Successor`	44,718	0	4,260
7	`Material Description`	44,718	0	4,426
8	`Detect Code Description`	44,718	0	9
9	`Source Of Cause Code Description`	44,718	0	16
10	`Material Notification Quality Sufficient`	44,718	0	1
11	`Notification Type`	44,718	0	1
12	`Plant For Material`	44,718	0	1
13	`Equipment Created Date`	44,718	0	3,151
14	`ABC-indicator`	40,326	4,392	3
15	`XYZ-indicator`	34,881	9,837	3
16	`Planned Delivery Time`	44,718	0	86
17	`Standard Price`	44,718	0	5,473
18	`LatestSuccessorDemand`	44,718	0	683
19	`RSR`	33,874	10,844	193
20	`Duration of Non-Defect`	44,718	0	3,168

Table 5.3: Summary of Cleaned Data Columns with Non-Null Counts, Null Counts, and Unique Counts

5.4.1. Explanatory Data Analysis

An analysis of ASML's complete dataset of defect notifications was conducted, aggregating all defects of all parts by month. This examination, which includes monthly data and seasonal decomposition, reveals crucial insights into the trends and factors influencing historical defect occurrences. The findings are essential for understanding and forecasting the return flows of defective parts. The graph in Figure 5.4 illustrates the trend and non-seasonal pattern.

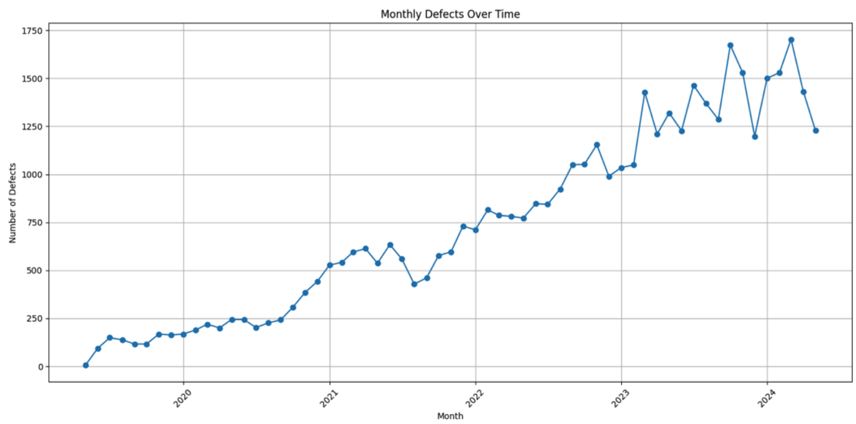


Figure 5.4: Total defects aggregated per month over the last 5 years (Author's Own Creation)

The data indicates a clear upward trend in defect notifications from 2019 to 2024, likely due to ASML's increased production volumes in the last years. As customer orders have risen, machine production has expanded, leading to a higher frequency of defects. This direct correlation between production scale and defect frequency underscores the impact of increased manufacturing activity on defect occurrences.

Additionally, ASML has significantly improved its data storage and management practices over the years, resulting in more reliable and accurate data, particularly in recent years. Historical issues such as missing values or inaccuracies have been addressed, enhancing the trustworthiness of recent defect counts and making them more reflective of actual conditions.

Seasonal decomposition of the defect counts indicates periodic fluctuations, suggesting some level of seasonality. However, these patterns are not dominant and likely result from operational cycles, maintenance schedules, or other periodic factors. While seasonality could be considered in forecasting models, it does not primarily drive defect counts in this dataset.

The residual component in the seasonal decomposition plot shows significant variability, indicating the presence of unexplained factors influencing defect counts. This variability could be due to unexpected events, changes in production processes, supply chain disruptions, or shifts in defect reporting practices, contributing to random fluctuations in the data.

5.4.2. Early Life Failure

Figure 5.5 illustrates the duration in days from equipment creation to defect notification, revealing that most defects occur within the first 1000 days of a part's lifecycle, particularly emphasizing the first 100 days. This pattern underscores the prevalence of early-life failures. The zoomed-in view highlights that despite some parts having longer lifespans due to repairs or upgrades, early-stage failures are predominant. This analysis supports the thesis focus on early-life failures and aligns with the problem statement, which emphasizes the severity of defects during the infant stage of the product lifecycle, particularly in ASML's Veldhoven factory. The data analysis corroborates the literature review, which points out the limitations of traditional reliability models for early-stage defects, thereby justifying the need for alternative approaches and the thesis's emphasis on this critical early stage.

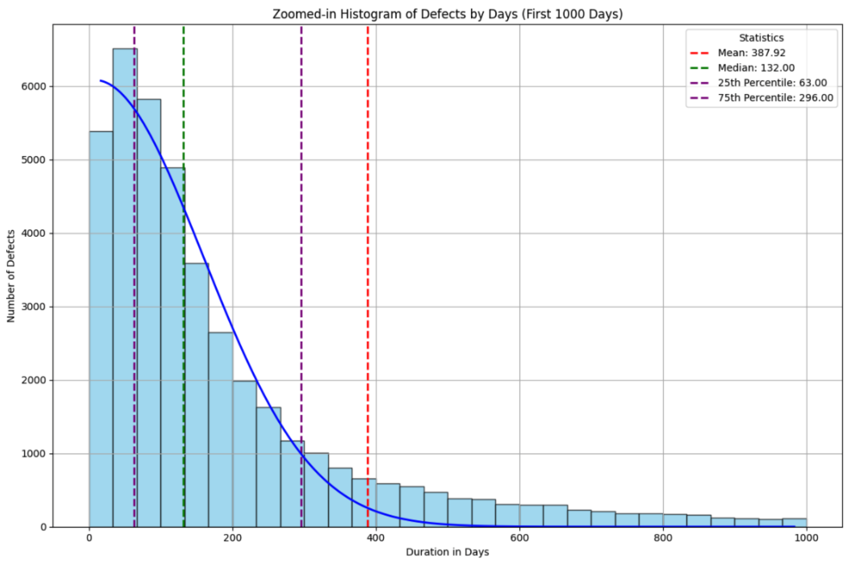


Figure 5.5: Early Life Defects Pareto (Author's Own Creation)

5.4.3. Detect & Cause of Failures

Two critical attributes used to identify the reasons for failures are the `Detect Code Descriptions` and the `Source Of Cause Code Descriptions`.

`Detect Code Descriptions` refer to how and when a failure was identified. This includes information on whether the part was defective upon arrival or if it failed during processing. There are nine distinct detection codes used to categorize these scenarios. The top five most frequently occurring detection codes are shown in Figure 5.6. Notably, one of these codes is labeled "Other" which is used when the exact reason cannot be determined from the code itself. In such cases, the specific details can often be found in the SAP system by referring to the MN description file. The majority of issues arise either from defective parts arriving from stock or failures occurring during processing.

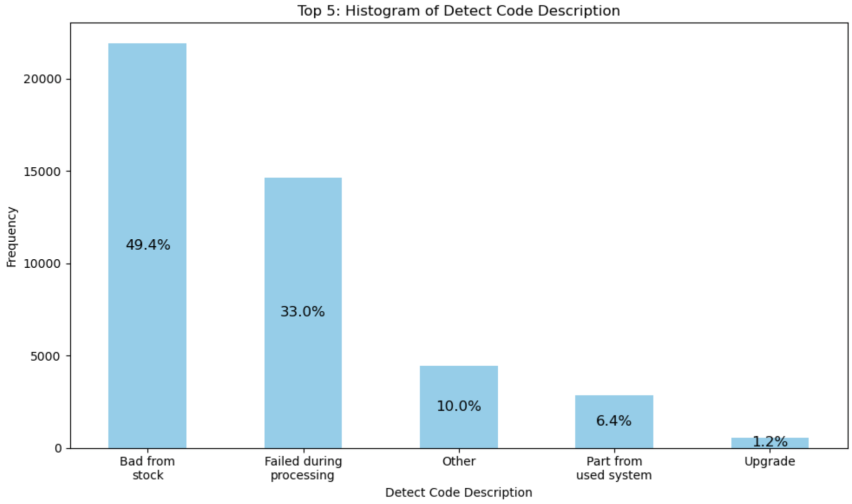


Figure 5.6: Top 5 Detection Code Descriptions (Author's Own Creation)

The `Source Of Cause Code Descriptions` are integral to the subsequent failure analysis phase. Engineers utilize these codes to identify the specific components and reasons behind failures. There are 16 different source of cause codes, and Figure 5.7 illustrates the five most common ones. Notably, two codes are related to vendor issues. One of these codes is used when vendor involvement is suspected, which triggers discussions between the vendor and engineers. If a RCA confirms the vendor's responsibility, a distinct code is assigned. Most failures occur during the production manufacturing process, particularly during assembly, and are likely due to quality or human errors. Technological issues, which rank as the fifth most common, differ from production issues in that they pertain to design failures of the parts rather than the assembly process itself.

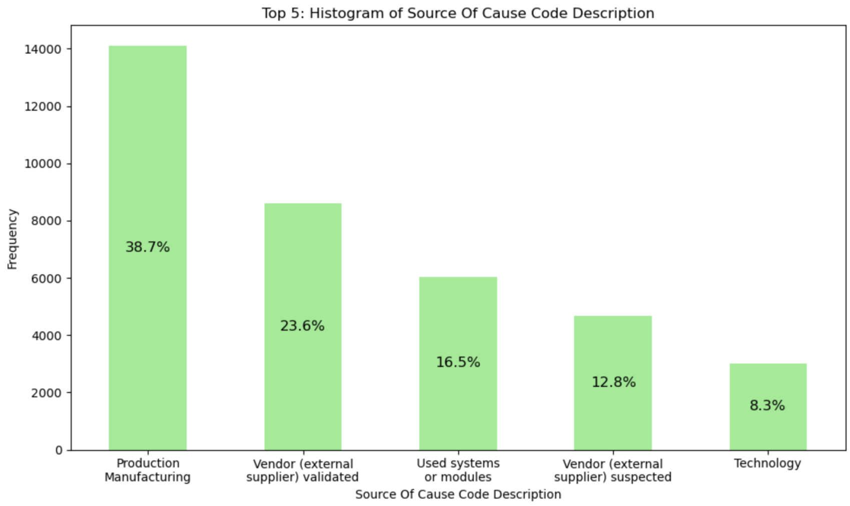


Figure 5.7: Top 5 Source of Cause Code Descriptions (Author's Own Creation)

5.5. Parts Classification

Classification is crucial for understanding and managing the return flow of defective parts from the factory. By categorizing parts based on defect patterns, defect count ranges, and economic impact, it becomes possible to systematically utilize the data in forecasting models and optimize future inventory management strategies according to the specific characteristics of different parts. This section explores various classification methods, starting with demand pattern classification and progressing to a tailored approach that combines traditional inventory analysis techniques.

Important Note: From this section onward, confidential data is used for analysis and classification. This data is also utilized in subsequent sections and chapters of this research. Due to privacy concerns and confidentiality agreements between the researcher, the university, and the company, actual material numbers (12NC) and corresponding characteristics, such as parts' standard prices, have been anonymized.

5.5.1. Demand Pattern Classification

To evaluate the behavior of the return flow and classify it, as well as to determine if there is a relationship between the return flow of defective parts and spare parts demand, the approach suggested by Syntetos and Boylan (2005) was initially applied. In this context, the ADI refers to the interval between notifications of defective parts, rather than the interval between demands. This ADI and CV^2 measures the return flow of defective parts, thus providing insight into the defect patterns rather than demand patterns. The results for the defective parts return flow from the ASML factory, based on the cleaned dataset, are presented in Table 5.4. These results show a significant similarity to the findings in spare parts demand forecasting as described by Lamghari-Idrissi (2021).

Table 5.4: Initial Classification Results

	Low ADI	**High ADI**
High CV^2	0.5%	91.2%
Low CV^2	0.0%	8.3%

The classification results indicate that 0.5% of the notifications fall into the ”Erratic” category and 8.3% into the ”Intermittent” category. Notably, these percentages are just outside the threshold lines, bordering the ”Lumpy” category. Figure 5.8 visually represents the classification of defective parts notifications and shows this threshold with dashed lines.

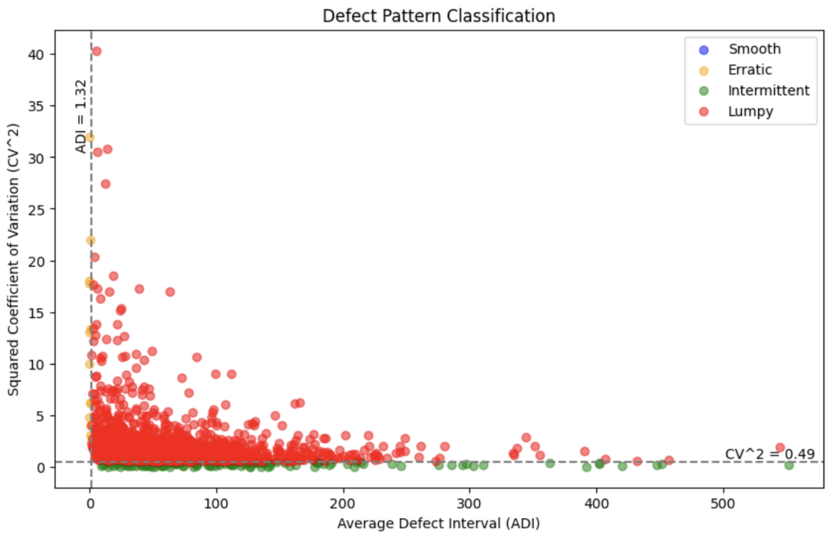


Figure 5.8: Defective Parts Notifications Pattern (Authors' own creation)

Given the predominance of the ”Lumpy” category in these research outcomes (91.2%), it becomes evident that while this classification is a useful starting point, additional tailored classification methods may be necessary to effectively address the unique characteristics of the dataset. This outcome highlights the similarities to spare parts demand forecasting and suggests that other classification analyses need to be considered to provide a comprehensive understanding of defect frequency patterns in semiconductor manufacturing.

5.5.2. ABCD-analysis

Traditional inventory management methods, such as ABC-analysis and FSN-analysis, optimize inventory control and are also used in demand forecasting by prioritizing parts based on annual usage value and frequency of parts movement, respectively. However, these methods are not suitable for forecasting defects in semiconductor manufacturing due to the irregular and variable nature of defect occurrences across numerous parts (12NC's). These methods are primarily designed for classifying inventory parts rather than defect counts.

To address these challenges, a novel approach named ABCD-analysis is proposed, specifically tailored and developed by the researcher for this study. This customization was necessary as existing methods did not adequately address the classification of defect types. The ABCD-analysis combines elements from ABC-analysis, which categorizes parts based on a Pareto distribution to prioritize them according to their economic value, and the FSN analysis, which evaluates the frequency of parts' movement. This new method categorizes 12NC's based on defect frequency ranges and cumulative defect counts, ensuring a balanced and representative selection for forecasting model application. Traditionally, the ABC-analysis consists of three categories; however, in this research, a fourth category "D" is introduced. This extension aligns with the precedent set by Syntetos et al. (Feb. 2009), who expanded the traditional approach for spare parts forecasting to include categories D, E, and F, as discussed in the literature review (Section 2.4.2). This tailored ABCD-analysis thus provides a more nuanced and effective classification system for defect forecasting.

The choice of ABCD-analysis is particularly relevant due to its adaptability and broad applicability, making it a valuable tool for other researchers facing similar challenges in return or defect frequency contexts. By focusing on defect count volumes in the return flow rather than stock volumes in inventory management, the ABCD-analysis provides a structured approach to managing and predicting defect occurrences. The outcomes of the ABCD-analysis are illustrated in Figure 5.9, which displays histograms for the total defects in each category alongside the Pareto cumulative defect count distribution. The 12NC with the highest defect counts in each specific ABCD category is shown at the top of the bars, with these identifiers anonymized for confidentiality.

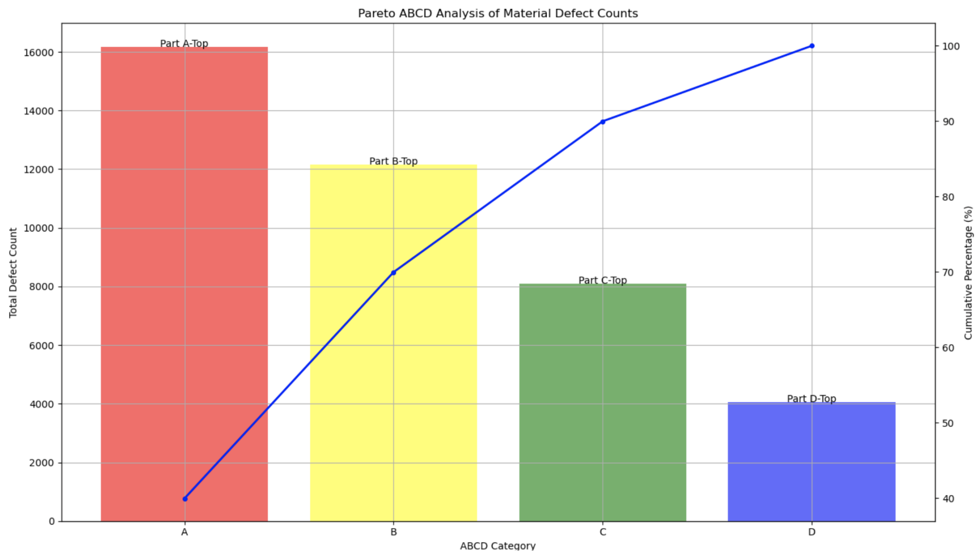


Figure 5.9: ABCD-analysis Histogram with Pareto (Authors' own creation)

To ensure a rational and impactful selection of parts, the economic value for each 12NC is calculated using the formula: $Economic\ Value = Demand \times Standard\ Price$. This calculation provides a justified basis for assessing the economic significance of each part. This evaluation is crucial for determining the business importance of the parts and ensuring that the selected components for analysis in the forecasting model are highly relevant. By adopting this method, other researchers can benefit from a comprehensive and adaptable framework for defect forecasting in various industries. Table 5.5 presents the ranges for demand, standard price, economic value, and defect count for each category. The values in this table have been normalized to ensure confidentiality and are not the actual figures.

The ABCD-analysis of defect counts and economic impact for 12NC parts reveals a critical subset of parts, with Category A parts representing a significant proportion of total defects. The selection criteria

Table 5.5: Ranges for Each Category (Normalized)

Cat.	Demand		Standard Price		Economic Value		Defect Count	
	min	*max*	*min*	*max*	*min*	*max*	*min*	*max*
A	0.0	5,233	0.05	0.50	0.0	2,616.5	39	782
B	0.0	4,774	0.03	0.80	0.0	3,819.2	15	38
C	0.0	14,048	0.04	0.75	0.0	10,536	5	14
D	0.0	4,948	0.02	1.00	0.0	4,948	1	4

for these 12NC parts are based on their variability in defect counts and economic value, allowing for an analysis of parts with diverse data points and economic impact. This study will concentrate on three parts from Category A due to its substantial impact and range of defect counts, spanning from 39 to 782. Additionally, the study will include two parts from Category B, which is the second most impactful category with defect counts ranging from 15 to 38. Only one part from Category C will be investigated due to its lower occurrence and defect counts, which range between 5 and 14. The exclusion of Category D is academically justified based on the low impact and unpredictability of defects in this category. Time series models are most effective when applied to data with identifiable patterns and sufficient frequency (Box et al., 2015). The random nature of defects in Category D does not lend itself to reliable forecasting, as the data lacks the consistency required for model training and validation. Concentrating on Categories A, B, and C allows for a more robust and meaningful analysis, aligning with the research objectives and operational priorities of ASML.

Table 5.6 summarizes the total defect count, percentage of total defects, unique 12NC's count, and percentage of total 12NC's for each category.

Table 5.6: ABCD-Categorically Aggregated Data

Cat.	Total Defect Count	% Total Defects	Unique 12NC's Count	% Total 12NC's
A	17,851	40%	235	5%
B	13,449	30%	591	14%
C	8,946	20%	1,131	27%
D	4,472	10%	2,303	54%
Total	39,129	100%	3,549	100%

- **Category A**: Comprising only 5% of the total 12NC parts, Category A is responsible for 40% of the total defect counts. These parts are critical for both defect reduction and cost management. Three 12NC parts with different defect count observations from Category A will be the primary subjects of this research.
- **Category B**: Representing 14% of the total 12NC parts, Category B accounts for 30% of the defect counts. Two 12NC parts with different defect count observations from Category B will be included in the study to provide a broader perspective on defect patterns.
- **Category C**: Representing 27% of the total 12NC parts, Category C accounts for 20% of the defect counts. One 12NC part from Category C will be analyzed to include an investigation of low-frequency defect dynamics, providing a comprehensive understanding of these less common defective parts.
- **Category D**: Including 54% of the total 12NC parts but accounting for only 10% of the defect counts. Due to the low and sporadic occurrence of defects, Category D parts will be excluded from this research, as forecasting models are less effective for such random and infrequent defect patterns (Box et al., 2015).

Table 5.7 highlights the top 6 highest parts in each category, focusing on the parts that should be the primary focus for detailed analysis. The parts are identified by their 12NC; however, for confidentiality reasons, they are labeled according to their category and ranked by frequency, as indicated by the number following the classification letter. The values in this table have been normalized to ensure confidentiality and are not the actual figures. This table introduces the "Impact Score", which is calculated by multiplying the "Economic Value" by the "Defect Count". This metric provides a more comprehensive measure that accounts for both the economic value and the frequency of defects. The "Impact Score" is particularly valuable from a research perspective as it incorporates defect frequency while maintaining a focus on economic impact, thereby aligning technical and business considerations.

Table 5.7: Top 6 Highest 12NC Parts per Category (Normalized)

Cat.	12NC Parts	Demand	Standard Price	Economic Value	Defect Count	Impact Score
A	Part A1	287.0	0.05	14.35	349	5,006.15
A	Part A2	75.0	0.20	15.00	230	3,450.00
A	Part A3	82.0	0.80	65.60	56	3,673.60
B	Part B1	15.0	1.00	15.00	36	540.00
B	Part B2	64.0	0.63	40.32	22	888.96
C	Part C	112.0	0.33	36.96	13	480.48

The correlation between the demand for parts and the defect counts suggests that parts with higher demand tend to exhibit higher defect counts. This correlation can be attributed to the larger volume of parts handled, where increased batch sizes may inherently lead to a higher probability of defects. However, this relationship varies for some parts, indicating the complexity of factors influencing defect occurrences.

Furthermore, the significant differences in defect counts across categories indicate a broad range of defect frequencies. This variability is beneficial for research purposes, as it allows for a diverse set of experiments with different input datasets in the forecasting models. Testing on parts with varying defect frequencies enhances the robustness and applicability of the forecasting models across different scenarios.

5.6. Classified Parts Analysis

The analysis begins by examining how defects occur over time for each unique 12NC part. Table 5.8 provides essential statistics for each unique part, including the first and last dates of defect occurrence, the total days between these dates, the mean duration between non-defective periods, and the mean duration between defect notifications. For example, Part A1 has an observation period of 1,321 days, frequent defects with a short mean interval of 3.80 days between notifications, and a mean non-defective period of 170.43 days. Part A2 has fewer defects but a significant mean non-defective duration of 460.63 days, with a mean interval of 2.66 days between notifications. Conversely, Part C shows the longest mean interval of 46.92 days between defects, indicating infrequent occurrences. These insights help in identifying parts with frequent issues and planning maintenance strategies effectively.

Table 5.8: Statistics for Each Unique Part

Cat.	12NC Parts	First MN Date	Last MN Date	Last – First Difference *days*	Mean Duration To Defect *days*	Mean Duration Between MN's *days*
A	Part A1	2020-10-12	2024-05-25	1,321	170.43	3.80
A	Part A2	2022-09-30	2024-05-31	609	460.63	2.66
A	Part A3	2021-12-06	2024-04-22	868	214.07	15.78
B	Part B1	2021-03-09	2024-03-04	1,091	73.72	31.17
B	Part B2	2023-03-30	2024-05-30	427	198.50	20.33
C	Part C	2022-10-17	2024-05-02	563	160.62	46.92

Figure 5.10 illustrates the distribution of defects over time for each unique part. The zero values are shown as points, but because there are predominantly zeros in the daily defect counts, the zeros appear as continuous lines on the x-axis, while the points greater than zero indicate actual defect occurrences. This visualization highlights the frequency of zero counts, demonstrating the sparsity of the data. The data's

lumpiness, characterized by periods of no defects followed by sudden bursts of multiple defects, mirrors the demand patterns often seen in spare parts.

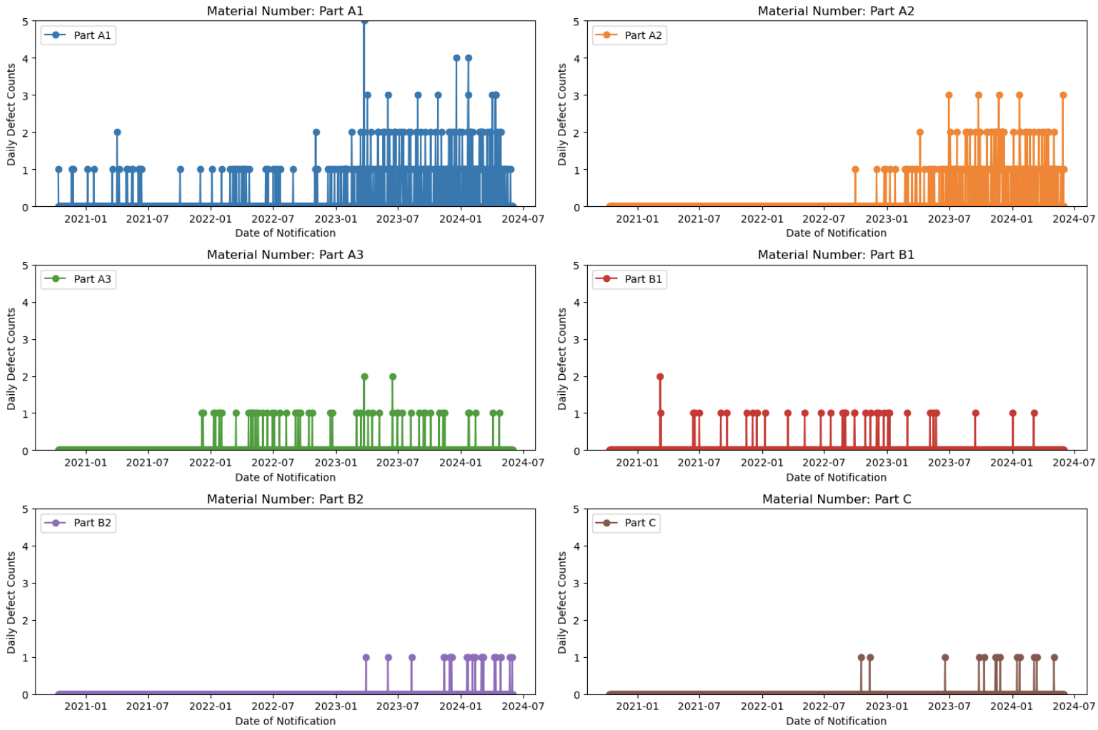


Figure 5.10: Unique Parts Defect Distribution (Authors' own creation)

Figure 5.11 displays the cumulative defect counts over time for each unique part. This visualization aids in evaluating the total number of defects accrued by each part throughout the observed period. It offers insights into the general occurrence of defects, revealing that nearly all parts experienced an increasing count of defects. For example, Part A1 demonstrates a significant rise in cumulative defects, suggesting a higher frequency of defect occurrences compared to other parts, which could potentially influence the performance of forecasting models.

5.7. Data Transformation

This section covers 'Data Resampling' and 'Data Split', explaining the methods used to enhance the dataset's suitability for forecasting models.

5.7.1. Data Resampling

The primary dataset for time series forecasting quantifies defects at specific moments through timestamps. Due to its sparse nature, with numerous zero values and lumpiness, data transformation is essential for enhancing suitability for forecasting models. This section outlines the transformation process using the `Pandas` library in Python, specifically through resampling techniques. Resampling aggregates data into larger time intervals, reducing noise and the prevalence of zero values. For this research, two types of resampled data will be considered for all six parts under study:

- **Weekly Resampled Data:** Data is aggregated into weekly intervals using Pandas' resampling functionality to always choose the last day of the week, Sunday, with 'W-SUN'. This reduces the number of zero values and smooths out daily fluctuations, providing a clearer view of defect trends. This frequency aligns well with ASML's RSC operations team's strategy, aiding in the prediction of defect occurrences for specific parts. The distribution of the resampled weekly data for all six parts is shown in Figure D.1 in the Appendix D. Although no significant seasonality is detected, some cyclical patterns are observed, particularly in Part A2 (see Figure 5.12).
- **Monthly Resampled Data:** Monthly forecasting is often preferred in literature due to its effectiveness

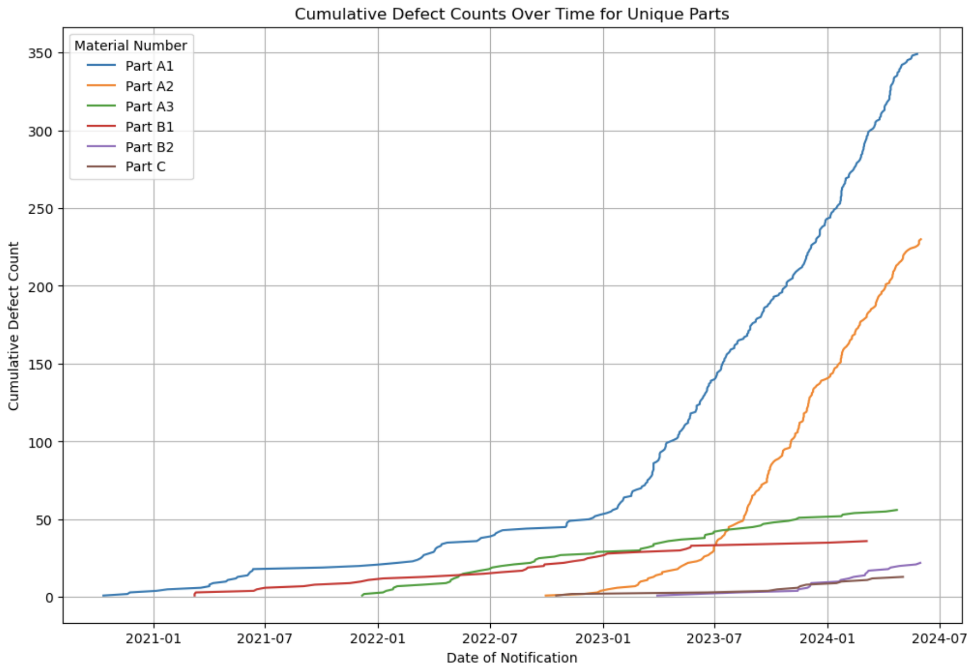


Figure 5.11: Unique Parts Cumulative Distribution (Authors' own creation)

in various studies. Additionally, forecasting on a monthly basis aligns with ASML's practice of updating KPI's and inventory numbers at the beginning of each month, ensuring the forecasts remain relevant and integrated with overall operational planning. The distribution of the resampled monthly data is shown in Figure D.2 in the Appendix D. Similar to the weekly data, no significant seasonality is detected, but some cyclical patterns are observed in Part A2 (see Figure 5.12).

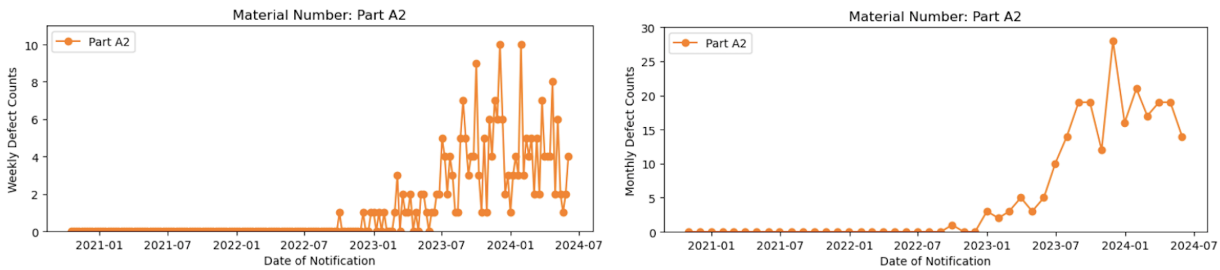


Figure 5.12: Part A2 Resampled Distribution (Left: Weekly; Right: Monthly)

By applying these resampling techniques, the transformed datasets (weekly and monthly) are expected to provide a more robust basis for forecasting models. The choice of resampling interval depends on balancing the operational needs of ASML's RSC operations team with empirical findings from literature, ensuring the forecasts are both practically and academically sound.

In conclusion, while no significant seasonality patterns were identified in the dataset, the presence of cyclical patterns, as shown in Part A2, will be considered in the forecasting models.

5.7.2. Data Split

In the context of this study, an ideal approach for data splitting would involve an annual division to capture full cycles of seasonal variations and trends. This method aligns with business objectives such as improved planning and promoting the reuse of existing parts, thereby reducing the necessity for new purchases. However, this approach is not feasible due to the sparse nature of the available data. Specifically, the six parts under investigation, along with most other parts in the same categories, have an average planned delivery time of 302 days. This near-annual delivery time, which begins when the purchase order is placed with vendors, prevents the complete blocking of new buys for the majority of parts in this case study.

Despite the limitation of not always being able to implement such a data split in scenarios with longer planned delivery times, it remains theoretically feasible for other research contexts where delivery times are shorter. In this study, the goal was to extract as many time steps as possible for each part to ensure effective model training. Literature commonly recommends an 80% training data and 20% test data split to achieve the most reliable forecasts, crucial for applications such as new buys blocking and capacity planning.

The 80/20 data split approach does more than just dividing the data; it also considers the period length for each split, ensuring that the model is trained and tested on data that is both comprehensive and representative of different temporal scenarios. The absolute length of the period used for the 80/20 split is modified to enhance the relevance and applicability of the forecasts.

The justification for the data split is as follows:

- **Sufficient Training Data:** Utilizing an 80% training data split guarantees that a substantial portion of the dataset is allocated for model training. This is crucial for enabling the models to learn and capture complex patterns and trends within the data, thereby improving their predictive accuracy. The long-term and short-term variations in defect occurrences can be better understood through this approach.
- **Robust Testing Set:** The 20% reserved for testing ensures a sufficiently large sample to evaluate the model's performance effectively. This helps in assessing the model's ability to generalize to unseen data, which is critical for the credibility and reliability of the forecasts.
- **Empirical Support:** The 80/20 split is empirically supported and widely used in various forecasting applications. Studies have shown that this ratio provides a balance between training the model adequately and having enough data to robustly test its performance. This split is well-documented in the literature, ensuring a robust methodological approach and facilitating comparability with existing studies.
- **Business Relevance:** The chosen split is highly relevant from a business perspective, supporting the management of incoming return flow and monthly operations. Short-term forecasting is essential for reverse inventory management, enabling the Reverse Supply Chain Operations (RSC Ops) team to efficiently plan and allocate resources for handling returns. The chronological aspect of the split means that the test period varies between 84-259 days for weekly data and 61-244 days for monthly data, thus covering both short- and medium-term forecasting horizons.
- **Absolute Length of the Period:** The period length for the 80/20 split is carefully selected to ensure that the historical data used is relevant to the forecasting model. Not every period has the same relevance; thus, the absolute length of the training and testing periods is adjusted to reflect the specific characteristics and requirements of each part. This consideration ensures that the models are trained on data that is temporally relevant, which is critical for achieving accurate and meaningful forecasts.

The weekly and monthly statistics for the 80/20 split are presented in Table 5.9 and Table 5.10, respectively. These tables provide detailed insights into the data distribution, including start and end dates and the duration of the test split in days. Differences in defect counts between weekly and monthly data highlight the granularity and varying alignment of weeks and months, which must be taken into account when interpreting the results.

To enhance clarity and readability, standardized abbreviations have been created for each part. These abbreviations will be used consistently throughout the document to simplify references and avoid redundancy. Table 5.11 defines these abbreviations.

Table 5.9: Weekly Statistics for 80/20 Split

Cat.	12NC Parts	Train Samples *defects / zeroes*	Test Samples *defects / zeroes*	Start Date	End Date	Test Split *days*
A	Part A1	176 / 78	173 / 0	2023-09-10	2024-05-26	259
A	Part A2	160 / 19	70 / 0	2024-02-04	2024-06-02	119
A	Part A3	49 / 61	7 / 19	2023-11-12	2024-04-28	168
B	Part B1	33 / 98	3 / 29	2023-08-06	2024-03-10	217
B	Part B2	15 / 38	7 / 8	2024-03-10	2024-06-02	84
C	Part C	8 / 57	5 / 12	2024-01-14	2024-05-05	112

Table 5.10: Monthly Statistics for 80/20 Split

Cat.	12NC Parts	Train Samples *defects / zeroes*	Test Samples *defects / zeroes*	Start Date	End Date	Test Split *days*
A	Part A1	175 / 9	174 / 0	2023-09-30	2024-05-31	244
A	Part A2	140 / 2	90 / 0	2024-01-31	2024-05-31	121
A	Part A3	49 / 3	7 / 2	2023-11-30	2024-04-30	152
B	Part B1	33 / 10	3 / 5	2023-08-31	2024-03-31	213
B	Part B2	13 / 5	9 / 0	2024-03-31	2024-05-31	61
C	Part C	10 / 9	3 / 2	2024-02-29	2024-05-31	92

Table 5.11: Part Abbreviations

Cat.	12NC Parts	Frequency	Test Prediction Steps	Part Abbreviation
A	Part A1	Weekly	38	**A1_Wk**
		Monthly	9	**A1_Mh**
A	Part A2	Weekly	18	**A2_Wk**
		Monthly	5	**A2_Mh**
A	Part A3	Weekly	25	**A3_Wk**
		Monthly	6	**A3_Mh**
B	Part B1	Weekly	32	**B1_Wk**
		Monthly	8	**B1_Mh**
B	Part B2	Weekly	13	**B2_Wk**
		Monthly	3	**B2_Mh**
C	Part C	Weekly	17	**C_Wk**
		Monthly	4	**C_Mh**

5.8. Conclusion

This section synthesizes the research outcomes from the system analysis as addressed in this chapter, specifically focusing on the fourth sub-question of this research:

Research Question 4

What data is available at ASML for predicting the return flow volume of defective parts, and how can this data be processed with a focus on the crucial variables?

The data available at ASML for forecasting the return flow volume of defective parts is extensive and multifaceted, primarily sourced from the SAP ECC system and processed using TIBCO Spotfire. Key datasets include ZLifecycle and CSAP Material Notification, which provide detailed records on defect occurrences, encompassing attributes such as 12NC's, plant locations, part descriptions, and timestamps for defect events.

To make this data suitable for forecasting, several processing steps are undertaken. Initially, data extraction from Spotfire involves filtering to focus on recent and high-quality MN entries, ensuring relevance and accuracy. The data is enriched by grouping parts by their latest 12NC successors and by creating additional features, such as the duration of defects, which measures the time interval between equipment creation and defect notification. Comprehensive data cleaning processes address missing values, duplicates, and obsolete records, resulting in a refined and reliable dataset. The validation of this dataset ensures its integrity and completeness, crucial for effective forecasting.

The crucial variable for forecasting is time series-based, determined by the occurrences of defects recorded in the 'Material Notification' (MN). By counting these defects for specific 12NC's based on the timestamps of the MN creation dates, this variable becomes essential for developing a time series forecast model, enabling the prediction of future defects at specific time steps and thereby determining the return flow from the factory to the RSC.

Parts classification further refines the data analysis. The classification of demand patterns reveals that the majority of parts fall into the "Lumpy" category, indicating irregular and infrequent defect occurrences. The ABCD-analysis categorizes parts based on defect frequency and economic value, identifying three parts from Category A, two from Category B, and one from Category C as focal points for further investigation. This classification highlights the variability in defect occurrences and the economic impact across different parts.

Resampling techniques transform the data into weekly and monthly intervals, reducing noise and enhancing the detection of trends and patterns. This processing step is vital for improving the dataset's suitability for time series forecasting models. For all six selected parts, two different data sets (weekly and monthly) are considered, resulting in 12 scenarios to investigate the differences in data variability and forecasting accuracy.

An ideal approach for data splitting would involve an annual division to capture full cycles of seasonal variations and trends. However, this approach is not feasible due to the sparse nature of the available data. Specifically, the six parts under investigation have an average planned delivery time of 302 days. This near-annual delivery time, which begins when the purchase order is placed with vendors, prevents the complete blocking of new buys for the majority of parts in this case study. Despite this limitation, the 80/20 data split is used to extract as many time steps as possible for each part, ensuring effective model training. This split is justified by its empirical support and business relevance, covering both short- and medium-term forecasting horizons. The absolute length of the period used for the 80/20 split is modified to enhance the relevance and applicability of the forecasts.

In conclusion, the data available at ASML, combined with thorough data processing and enrichment, provides a solid foundation for forecasting the return flow volume of defective parts. The processed dataset, enriched with crucial variables and cleaned to ensure reliability, is well-suited for developing accurate time series forecasting models. The classification and resampling methods applied ensure that the data is effectively structured to capture the key trends and patterns necessary for robust forecasting. This approach not only enhances the relevance and accuracy of the forecasts but also tailors them to the unique operational context of ASML's RSC management, ultimately answering the fourth research question comprehensively.

6

Forecasting Models for Defective Machinery Parts

This chapter investigates various forecasting models to identify the most effective approach for predicting the return flow of defective machinery parts at ASML. The analysis begins by establishing a clear set of requirements for the selection of forecasting models, ensuring they address the unique challenges and complexities of ASML's operations. It then discusses variable selection and feature engineering, focusing on the key attributes necessary for accurate forecasting. The chapter proceeds to explore both univariate and multivariate models, detailing their selection, application, and the criteria used for their evaluation. Finally, the chosen evaluation metrics will be discussed. The chapter concludes by answering the fifth research question:

5. Which forecasting models and evaluation metrics are best suited for predicting the return flow of defective machinery parts at ASML, considering the specific requirements and crucial data variables?

6.1. Requirements

To choose the best-suited forecasting models for predicting the return flow of defective machinery parts at ASML, a clear set of requirements has been established. These requirements ensure that the selected models can effectively address the unique challenges and complexities of the semiconductor industry, particularly those relevant to ASML's operations. Given the lack of literature directly addressing the quantity return flow of defective machinery parts while accounting for various types of parts with varying amounts of data, these requirements are crucial for selecting models that can handle the intricacies specific to ASML's needs. As concluded in Chapter 5, time series forecasting is preferred for this research because the defect counts on specific dates for specific parts need to be determined to predict the return flow in RSC from the factory.

1. **Literature-Based Model Selection**: The forecasting model should be one that is well-documented in the literature, with a preference for models used to predict spare part demand patterns.
2. **Time Series Approach**: The model should be based on time series analysis to effectively utilize the available data from Chapter 5.
3. **Handling Non-Stationarity and Trends**: The model must be capable of managing non-stationary data and identifying trends within the dataset.
4. **Time Horizon Flexibility**: The model should be mainly suitable for medium-term forecasting horizons to predict defects one year ahead to block new buys, but also performance for short-term predictions could be valuable for managing the return inflow as indicated in Section 5.7.2.
5. **Balancing Complexity and Performance**: Considering the variety of parts in the semiconductor industry, a less complex model is preferred if it can provide accurate predictions. However, more complex models should also be considered if they offer significantly better performance, especially given potential future improvements in computational efficiency.
6. **Variable Data Volumes**: The model should be capable of handling different volumes of data, reflecting the diverse range of parts and their availability in the semiconductor industry.
7. **Performance with Sparse and Lumpy Data**: The model must perform well with sparse and lumpy data, which characterizes the majority (98.3%) of the data.

6.2. Variables Selection & Feature Engineering

The attribute columns described in Chapter 5 and summarized in Table 5.3 are not all relevant for time series forecasting models for specific parts. Some attributes are redundant as they remain constant for unique parts. However, during data analysis, these attributes were essential for characterizing and filtering the data to ensure quality. Certain attributes need to be retained for part identification, while others will serve as predictor variables in the forecasting models.

6.2.1. Univariate Models

For univariate time series models, the primary part identification attribute necessary for the forecasting models is `Material Number Clean 12NC Latest Successor`, which aggregates defect counts over time. The key temporal variable for these models is the `Date of Notification`, representing the timestamp of each defect notification (MN).

To enable potential traceability and further analysis within the system, unique identifiers such as `Material Number`, `Equipment Number cSAP`, and `Notification Number` are retained as attributes for the purpose of reverse analysis.

6.2.2. Multivariate Models

Multivariate models can enhance forecasting performance by incorporating additional exogenous dynamic variables specific to each `Equipment Number cSAP` and `Notification Number` within the cumulative `Material Number Clean 12NC Latest Successor`. Feature engineering plays a crucial role in these models by decomposing the date into multiple components to capture cyclical and seasonal patterns.

Other exogenous features, such as temperature, are not relevant to this research because the parts are used in clean rooms regulated for specific temperature and humidity levels, making outside weather factors irrelevant. Furthermore, future orders and sales data were not available for this study, and therefore, could not be incorporated into the forecasts. The date-related features utilized in this research include:

- `Year`: The calendar year associated with the `Date of Notification`.
- `Month`: The numerical representation of the month for the `Date of Notification`.
- `Week`: The specific week number within the year corresponding to the `Date of Notification`, used only for the weekly sampled data.
- `Holiday in Month`: A binary variable indicating whether the `Date of Notification` falls on a public holiday in the month, used only for the monthly resampled data.
- `Holiday in Week`: A binary variable indicating whether the `Date of Notification` falls on a public holiday in the week, used only for the weekly resampled data.

However, certain dynamic variables specific to each `Equipment Number cSAP` and `Notification Number` within the cumulative `Material Number Clean 12NC Latest Successor` will not be incorporated. These variables, which include:

- `Detect Code Description`
- `Source of Cause Code Description`
- `Duration of Non-Defect`

will not be taken into account because only historical data is available for these variables. For time series forecasting, it is essential to have future values of these variables to use them in the predictions. Without this future data, the models are unable to function correctly. Consequently, while these features could be utilized in future research through regression models to forecast their influence on defect counts, they will not be employed in the experiments of this research.

6.3. Selection of Forecasting Models

The forecasting models in Chapter 3 are compared to each other in the Figure 3.4 to these different requirements which are converted to characteristics which were discussed in the various literature.

Given the specific requirements and complexities inherent to the semiconductor industry's return flow of defective machinery parts at ASML, a selection of forecasting models is made to effectively address these challenges. This section details the chosen models, justifies their selection, and explains their application within this research framework. The base forecasting models selected for this study, along with their relationships to the research, are highlighted in blue in the summarized literature matrix shown in Figure 6.1.

		General & Sector Specific						Forecasting Characteristics								
Forecasting Models	References	Logistics / SC Usage	Return Flow	Failure / Defect of Parts	Spare Parts Demand Pattern	Sales Input Variable	Complex Patterns	Time Series Based	Multivariate	Non-Stationary	Trend	Seasonality	Time Horizon (Short, Medium, Long)	Complexity	Data Volume Required	Performence for Large Variation Quantity & Time
SES	1,5,8	●	○	○	○	●	○	●	○	○	○	○	S & M	+	+	–
ARIMA	1,10,13,14,15	●	●	●	●	●	●	●	○	●	●	○	S & M	– / +	– / +	+
ARIMAX	12	●	●	○	○	○	●	●	●	●	●	○	M	–	– / +	+
LSTM	32,33,34,35	●	●	●	●	●	●	●	●	●	●	●	M & L	–	–	+
This Research		●	●	●	●	○	●	●	●	●	●	○	S & M	–	– & +	+

Figure 6.1: Selected Models Matrix from Literature Research (Authors' own creation)

6.3.1. Simple Exponential Smoothing

SES is chosen as the initial benchmark model due to its simplicity and ease of use. It provides a straightforward approach to smoothing data, making it an ideal foundational model for forecasting. The α parameter in SES will be optimized to determine the best forecast. SES is well-documented in the literature for basic time series forecasting, particularly for smoothing data and providing a clear baseline. As a simple model, it serves as a baseline, allowing for the comparison of the performance of more advanced models. By starting with a simple model like SES, subsequent enhancements and modifications can be evaluated for their added value. This model satisfies the requirement for a literature-based, time series approach while offering a clear starting point for evaluating more complex models. However, it has limitations in handling trends and complex patterns in non-stationary data.

6.3.2. ARIMA

The ARIMA model is chosen for forecasting the return flow of defective machinery parts at ASML due to its strong alignment with the specific requirements of this research. ARIMA is extensively documented in the literature for its effectiveness in forecasting spare parts demand and failure rates, ensuring its reliability within ASML's context.

The ARIMA model is particularly well-suited for non-stationary data, which allows it to identify and model trends within the return flow data. This capability is crucial for handling the variability and potential lumpiness observed in ASML's data. The model's robustness in capturing the dynamics of return flow data, especially with sparse datasets, makes it an ideal choice.

The model's parameters (p, d, and q) will be optimized using Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) plots. Typically, the objective is to minimize the AIC, ensuring an efficient and accurate fit that balances complexity and performance. However, in this research, the model will be optimized to achieve the lowest possible evaluation metrics, which will be determined in further analysis. This approach ensures that the model's performance can be compared generally and provides the most accurate predictions for ASML's needs.

Additionally, ARIMA's flexibility allows it to handle varying data volumes, which is essential for the diverse range of parts at ASML. This combination of attributes—robustness, flexibility, and the ability to accurately capture trends in non-stationary data—makes ARIMA a robust and flexible choice, satisfying all the outlined requirements.

6.3.3. ARIMAX

The ARIMAX model is selected to enhance the capabilities of the base ARIMA model by incorporating exogenous variables, aiming to improve forecasting accuracy for ASML's return flow of defective machinery parts. The literature discusses the ARIMAX model for its superior forecasting performance in contexts where external variables are critical. Studies indicate that incorporating exogenous predictors can lead to more accurate forecasts by accounting for influences outside the primary time series data. This model's selection is also driven by the opportunity to directly compare its performance with the ARIMA model, as the literature suggests potential improvements but lacks specific testing in the RSC context.

The ARIMAX model extends the ARIMA framework by integrating external factors into the time series analysis, making it a multivariate approach. The exogenous variables provide deeper insights into defect patterns, allowing for a more comprehensive understanding of the factors affecting return flows. This enhanced model is expected to better capture the complexities and variability in ASML's defect data, providing more accurate and actionable forecasts.

The parameters of the ARIMAX model will be optimized using techniques such as ACF and PACF plots, along with minimizing evaluation metrics to ensure a precise and efficient fit. While the AIC will be checked, it will not be the primary criterion for optimization. This approach ensures that the model's performance is balanced with its complexity, achieving superior accuracy without imposing unnecessary computational burden. Similar to the ARIMA model it satisfies all the outlined requirements.

6.3.4. LSTM

The LSTM model is selected due to its exceptional capability to handle time series data with complex temporal dependencies. LSTM model is particularly effective in scenarios where long-term dependencies in the data are critical for accurate forecasting. This makes them highly relevant for predicting the return flow of defective parts in the semiconductor industry.

LSTM models excel in managing short to long-term dependencies and recognizing complex patterns, which are typical in forecasting spare parts demand. They are especially suited for scenarios where past events significantly influence future outcomes, such as parts failing during initial assembly stages in ASML's operations. The literature supports the superior performance of LSTM models compared to traditional methods like ARIMA, SVM, Random Forest, and ANN, particularly in scenarios involving sparse data and complex dependencies.

The LSTM model's architecture includes memory cells that can maintain information over long periods, allowing the model to learn and remember important features from the data. This ability is crucial for capturing the intricate dynamics of return flows in the semiconductor industry.

In the implementation of LSTM models, the focus will be on tuning hyperparameters such as the number of layers, the number of neurons per layer, the learning rate, and the batch size to optimize model performance. The training process involves backpropagation through time which updates the model weights to minimize the loss function.

In conclusion, the LSTM model is selected for its robust performance in handling complex temporal dependencies and its extensive support in the literature for forecasting scenarios with sparse and intermittent data patterns. By leveraging the LSTM model's advanced architecture, this research aims to achieve a higher accuracy in predicting the return flow of defective machinery parts compared to traditional forecasting methods, thereby satisfying all the outlined requirements.

6.4. Evaluation Metrics

Accurately assessing the performance of forecasting models requires selecting appropriate evaluation metrics that can handle the unique challenges of the dataset, such as the presence of zeros and high variability. This section outlines the selection, normalization, and ranking of evaluation metrics to ensure robust and fair comparisons of model performance.

6.4.1. Selection of Evaluation Metrics

Given the challenges posed by the presence of zeros in the data and high variability, the selection of evaluation metrics is crucial for accurately assessing the performance of forecasting models. The three most effective metrics for this research, based on the specific requirements and characteristics of the data, are presented in Table 6.1. These metrics are also commonly used in similar research from the literature.

Table 6.1: Selected Evaluation Metrics

Metric & Equation	Relevance to Research
$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	MSE is useful for highlighting significant forecasting errors due to the squaring of each error term, making it sensitive to large deviations. This characteristic is beneficial for assessing model performance in scenarios with high variability, which is evident from the available data in this research. MSE is used in similar research from the literature for models such as SES, ARIMA, and LSTM.
$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	MAE is chosen for its straightforwardness and robustness against outliers. It is indispensable for comparing series measured on the same scale and is simple to calculate and interpret. MAE provides an accurate measure of forecast accuracy by ensuring that negative and positive errors do not cancel each other out. MAE is used in similar research from the literature for models such as ARIMA and LSTM.
$mMAPE = 100\% \times \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\lvert\hat{y}_i - y_i\rvert}{1+\lvert y_i\rvert}\right)$	The mMAPE addresses the limitations of standard MAPE, particularly in datasets with zero or near-zero values. By ensuring the denominator is never zero, it enhances robustness and interpretability. This metric is valuable for evaluating models in the context of sparse and lumpy data, which is a characteristic of the available data in this research. mMAPE is used in similar research from the literature for the ARIMA and LSTM models.

6.4.2. Normalizing and Ranking the Metrics

To compare the performance of the forecasting models using different evaluation metrics, the values of MSE, MAE, and mMAPE are normalized to ensure they are on a comparable scale. This approach, similar to the one used by Lee et al. (2021), ensures that each metric contributes equally to the overall performance ranking. This normalization is achieved using the min-max normalization technique:

$$\text{Normalized Metric} = \frac{\text{Metric} - \text{min}(\text{Metric})}{\text{max}(\text{Metric}) - \text{min}(\text{Metric})}$$

Normalization ensures that each metric contributes equally to the overall performance ranking.

6.4.3. Aggregating Scores and Ranking Models

Following the approach used by Lee et al. (2021), after normalizing the metrics, an aggregate score for each model is calculated by averaging its normalized MSE, MAE, and mMAPE scores:

$$\text{Aggregate Score} = \frac{\text{Normalized MSE} + \text{Normalized MAE} + \text{Normalized mMAPE}}{3}$$

Models are then ranked based on their aggregate scores, with the model having the lowest aggregate score being considered the best performing. This process is repeated for each dataset frequency and part classification to identify the top-performing model in each scenario.

To determine the overall best performing model across all dataset frequencies and part classifications, the mean of the aggregate scores for each model across all scenarios is computed:

$$\text{Overall Score} = \frac{\sum_{i=1}^{N} \text{Aggregate Score}_i}{N}$$

where N represents the total number of different dataset frequencies and part classifications. This comprehensive ranking approach ensures a robust comparison and selection of the optimal parameters for each forecasting model. Additionally, it facilitates the identification of the best overall forecasting model for the given research context in subsequent analyses.

6.5. Conclusion

This chapter addressed the fifth sub-question of the research:

Research Question 5

Which forecasting models and evaluation metrics are best suited for predicting the return flow of defective machinery parts at ASML, considering the specific requirements and crucial data variables?

To determine the most appropriate forecasting models and evaluation metrics for predicting the return flow of defective machinery parts at ASML, the following models were carefully selected based on the stated requirements:

- **SES**: Chosen for its simplicity and as a benchmark for comparison. While SES does not meet all the requirements, particularly in handling trends and complex patterns in non-stationary data, it provides a foundational baseline to evaluate the performance of more complex models.
- **ARIMA**: Ideal for handling non-stationary data, ARIMA is capable of identifying and modeling trends within the return flow data, making it highly suitable for the variability and lumpiness observed in ASML's data. ARIMA meets all the specified requirements, ensuring reliability and accuracy.
- **ARIMAX**: Building on the ARIMA model, ARIMAX incorporates exogenous variables to improve forecasting accuracy. This allows for deeper insights into defect patterns by considering external influences. ARIMAX meets all the requirements, offering enhanced forecasting performance.
- **LSTM**: Selected for its ability to manage complex temporal dependencies, LSTM models are particularly effective for scenarios with sparse and intermittent data patterns, offering potentially higher accuracy in forecasting the return flow of defective parts. LSTM meets all the requirements, providing robust and advanced forecasting capabilities.

The selected evaluation metrics—**MSE**, **MAE**, and **mMAPE**—ensure a robust assessment of model performance. MSE is particularly useful for highlighting significant forecasting errors, MAE provides a straightforward measure of accuracy without being unduly influenced by outliers, and mMAPE addresses the limitations of standard MAPE by being robust to zero or near-zero values.

By normalizing and ranking these metrics, the models can be effectively compared, facilitating the identification of the optimal forecasting approach. This comprehensive evaluation ensures that the chosen models not only meet the specific requirements of ASML but also provide actionable insights for reverse inventory planning within the RSC.

7

Experiments & Results

This chapter builds on the extracted and transformed data from Chapter 5 and the selected models from Chapter 6, setting the foundation for the experiments. The chapter is organized into sections for each of the four forecasting models: SES, ARIMA, ARIMAX, and LSTM. Each section details the model configuration, parameter optimization, and performance results for all parts and frequencies. Figure 7.1 provides an overview of this chapter, which shows the data extraction, transformation, and connection with the experiments and results. Insights are provided into how these models handle different defect part classifications and data frequencies. Additionally, there is a section analyzing the computational effort required for each model. The chapter concludes by synthesizing the findings to identify the optimal parameters for each model, thereby addressing the sixth sub-research question:

6. What are the optimal parameters for the most suitable forecasting models tailored to different classifications of defective parts and varying data frequencies?

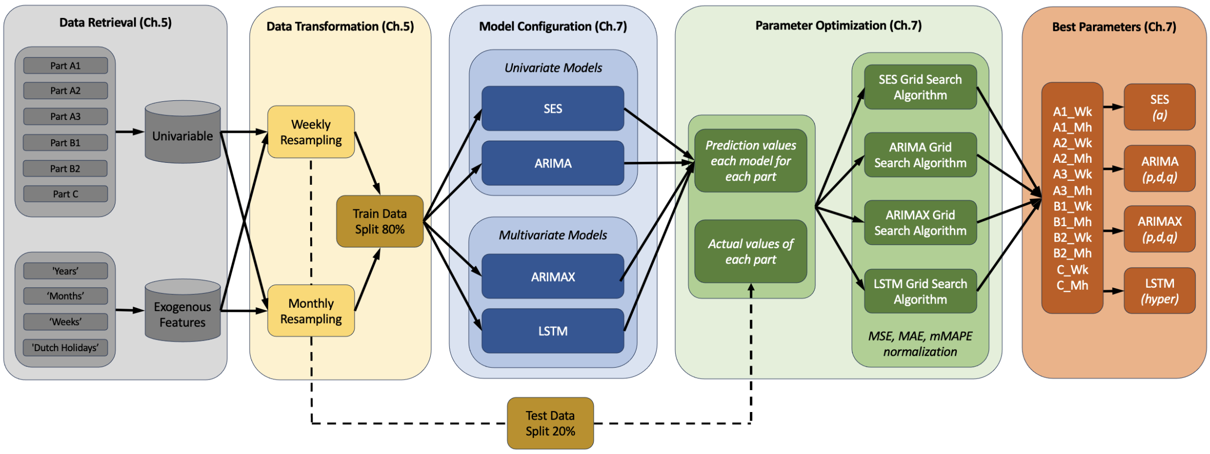


Figure 7.1: Data Extraction & Transformation connection with Experiments & Results

7.1. Simple Exponential Smoothing

SES is a foundational time series forecasting method used to predict future values based on past observations. This technique is particularly useful for data without trends or seasonal patterns. SES applies exponentially decreasing weights to past data, giving more importance to recent observations.

7.1.1. Model Implementation

SES forecasts are generated using the following formula:

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha)\hat{y}_{t|t-1}$$

In this equation, $\hat{y}_{t+1|t}$ represents the forecasted value at time $t + 1$ based on data up to time t. The parameter α is the smoothing constant, which ranges between 0 and 1, determining the weight given to the most recent observation y_t versus the previous forecast $\hat{y}_{t|t-1}$. The higher the value of α, the more weight is placed on the most recent observation.

The SES model was implemented in Python using the `statsmodels.tsa.holtwinters.SimpleExpSmoothing` library due to its efficiency and computational robustness.

7.1.2. Parameter Optimization

To determine the best α value for the SES model, a grid search algorithm is developed, as outlined in algorithm 1 which can be found in Section C.1. This algorithm systematically tests α values from 0.0 to 1.0 in increments of 0.025, ensuring a thorough exploration of the parameter space. The optimization process involves:

1. Fitting the SES model on the training data for each α value.
2. Forecasting defect counts on the test data.
3. Applying a non-negativity constraint to ensure predictions are non-negative.
4. Calculating evaluation metrics MSE, MAE, and mMAPE.
5. Normalizing these metrics to define comparison.
6. Identifying the best α based on an aggregate score of normalized metrics.

7.1.3. Results

The optimal α values and their corresponding performance metrics for each defect part are summarized in Table 7.1. The results are ranked by normalized MSE, MAE, and mMAPE across twelve different outputs. The rankings are categorized into groups A, B, and C, separated by dashed lines, with the models within each group ordered by their performance.

Table 7.1: Optimal Parameters and Performance Metrics for SES

Part_Freq	α	Avg. $\hat{y}$	MSE	MAE	mMAPE	Rank
A2_Mh	0.55	18.65	6.03	1.87	10.89	3
A3_Wk	0.85	0.13	0.31	0.35	20.95	7
A1_Wk	0.40	3.99	5.62	1.71	31.55	9
A2_Wk	0.05	3.74	3.45	1.50	36.38	8
A3_Mh	1.00	2.00	1.50	0.84	75.00	11
A1_Mh	0.525	17.45	46.7	5.74	34.28	12
B1_Wk	1.00	0.00	0.10	0.10	4.69	4
B2_Mh	0.575	2.09	1.51	0.98	21.33	5
B2_Wk	0.00	0.31	0.61	0.61	35.53	6
B1_Mh	1.00	0.00	0.38	0.38	18.75	10
C_Mh	0.00	0.63	0.71	0.75	47.40	1
C_Wk	1.00	0.00	0.30	0.30	14.71	2

Figure 7.2 shows the SES forecasts compared to actual defect counts for Parts C and A1, and additional plots for the other parts are available in Appendix D: Figure D.3. For Part C_Mh, the SES model performs well, with the predictions closely aligning with the actual counts. However, for the weekly data, it shows a constant zero prediction, indicating that the model performs well in predicting non-defective periods but fails to capture the occasional defects. This suggests that SES can handle parts with infrequent and low defect counts effectively on a monthly basis, but for weekly frequencies, its performance is misleadingly good due to the predominance of non-defective periods.

Notably, the predicted mean value ($\hat{y}$) for parts A3_Wk, B1_Wk, B2_Wk, and C_Wk is also zero or almost zero, suggesting that these models predict no defects. While this results in excellent performance metrics (very low MSE and MAE), it indicates that the models similarly are forecasting non-defects, which may not reflect the real-world scenario where some defects are expected. Thus, from a practical perspective, these models may not be as effective despite their seemingly good performance.

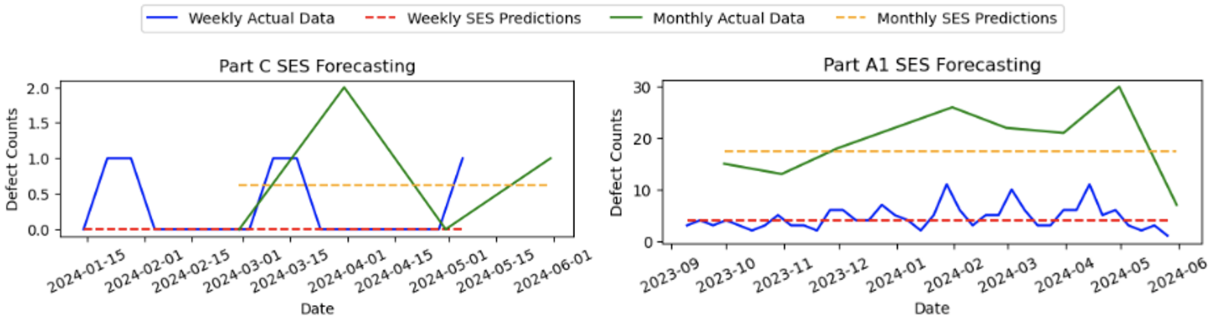


Figure 7.2: Weekly and Monthly SES Results for Part C (rank 1 & 2) and Part A1 (rank 9 & 12)

On the other hand, for both weekly and monthly Parts A1, the SES model is not able to accurately forecast the more variable and higher defect counts, as evidenced by the larger discrepancies between the predicted and actual values. The variability and higher frequency of defects in Parts A1 expose the limitations of SES in handling such complex patterns.

These results demonstrate the performance of SES in forecasting defect counts, providing valuable insights for further analysis. However, the limitations observed suggest the need for more advanced models to handle parts with greater variability and complexity in defect counts.

7.2. ARIMA Model

The ARIMA model is a widely-used statistical method for time series forecasting. It integrates AR, I, and MA components to capture various patterns in time series data. This section focuses on the basic ARIMA model without seasonality or exogenous variables.

7.2.1. Model Configuration

The ARIMA model is represented by the equation:

$$y_i = \sum_{j=1}^{p} \phi_j y_{i-j} + \epsilon_i + \sum_{j=1}^{q} \theta_j \epsilon_{i-j}$$

In its expanded form:

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - B)^d y_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q)\epsilon_t$$

where:

- ϕ and θ denote the coefficients for the AR and MA terms, respectively,
- d indicates the differencing order,
- B is the backshift operator,
- ϵ_t represents white noise.

The components of the ARIMA model are configured as follows:

- **Autoregressive (AR) Component**: Predicts the current value based on past values. The order (p) shows how many past values are used. The PACF plot helps identify p by highlighting significant lags.
- **Moving Average (MA) Component**: Predicts the current value using past forecast errors. The order (q) shows how many past errors are used. The ACF plot helps identify q by highlighting significant lags.
- **Integrated (I) Component**: Makes the data stationary by differencing. The order (d) shows how many times differencing is needed. The Augmented Dickey-Fuller (ADF) test helps determine d, ensuring the series becomes stationary.

By configuring these parameters, the ARIMA model can be tailored to the specific characteristics of the time series data. The model fitting process involves estimating these parameters to minimize prediction error, providing a robust tool for forecasting defect counts in a time-dependent manner.

In Python, the `statsmodels` library provides tools to implement ARIMA models. The primary functions used are:

- `ARIMA()`: Defines the ARIMA model with parameters specifying the order of AR (p), differencing (d), and MA (q) components.
- `fit()`: Estimates the parameters of the ARIMA model by fitting it to the provided time series data.

The `statsmodels` library is chosen for its comprehensive and flexible implementation of ARIMA models. It provides robust methods for parameter estimation, diagnostics, and validation, ensuring accurate and reliable forecasts. The library's functions allow for easy configuration and fine-tuning of model parameters, making it a preferred choice for time series analysis in both academic and practical applications.

7.2.2. Parameter Range Selection

Selecting appropriate parameter ranges for the ARIMA model is crucial for ensuring accurate and reliable forecasting. This process involves a detailed analysis of the ACF, PACF, and ADF test results to determine suitable values for p, d, and q.

The ACF and PACF plots for the original defect counts provide insights into the autocorrelation structure of the time series. These plots help identify potential values for the AR (p) and MA (q) parameters. For instance, the ACF plot for the original defect counts of Part A1_Wk shows significant positive autocorrelations up to several lags, suggesting that MA terms up to $q = 5$ should be considered. Similarly, the PACF plot displays significant spikes at initial lags, indicating potential AR terms up to $p = 14$.

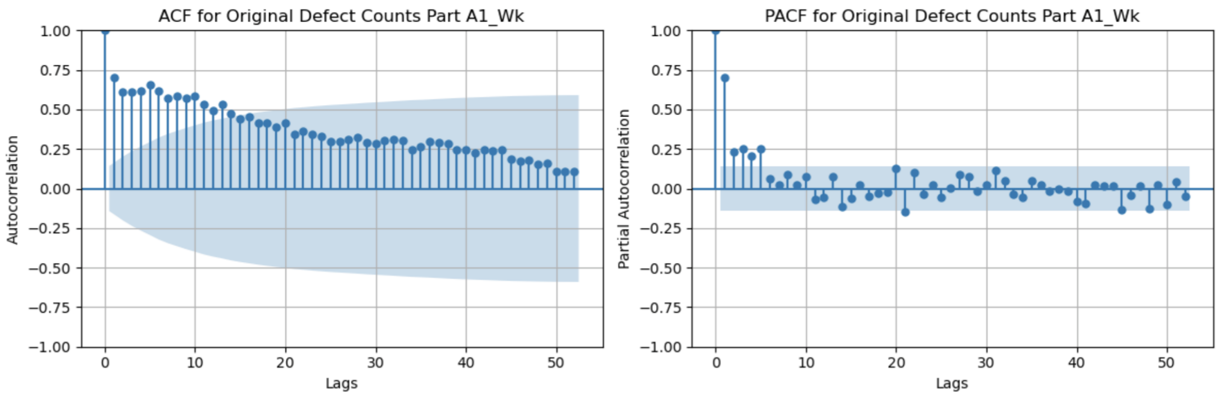


Figure 7.3: ACF and PACF plots for original defect counts of Part A1_Wk

The ADF test is used to assess the stationarity of a time series. For the Part A1_Wk dataset, the ADF test results for the original series revealed an ADF statistic of -1.41 and a p-value of 0.57. Since the p-value is greater than 0.05, the null hypothesis cannot be rejected, indicating that the series is non-stationary and necessitates differencing.

After applying first-order differencing ($d = 1$), the ADF test was conducted again. This time, the first differenced series yielded an ADF statistic of -6.76 and a p-value of 0.00, which is significantly less than 0.05. These results indicate that the null hypothesis can be rejected, confirming that the differenced series is stationary.

After differencing, the ACF and PACF plots were analyzed again to determine the appropriate values for p and q. The reanalysis for the differenced series showed reduced autocorrelations, with significant spikes indicating that p and q values up to 3 should still be considered. The ADF test for the differenced series indicated stationarity with a p-value less than 0.05.

Following similar analyses for all parts, as illustrated in Figure D.7 through Figure D.18 in Appendix D, the parameter ranges for grid search optimization were determined and summarized in Table 7.2. This comprehensive approach ensures that the selected ranges encompass all potential values while maintaining computational feasibility.

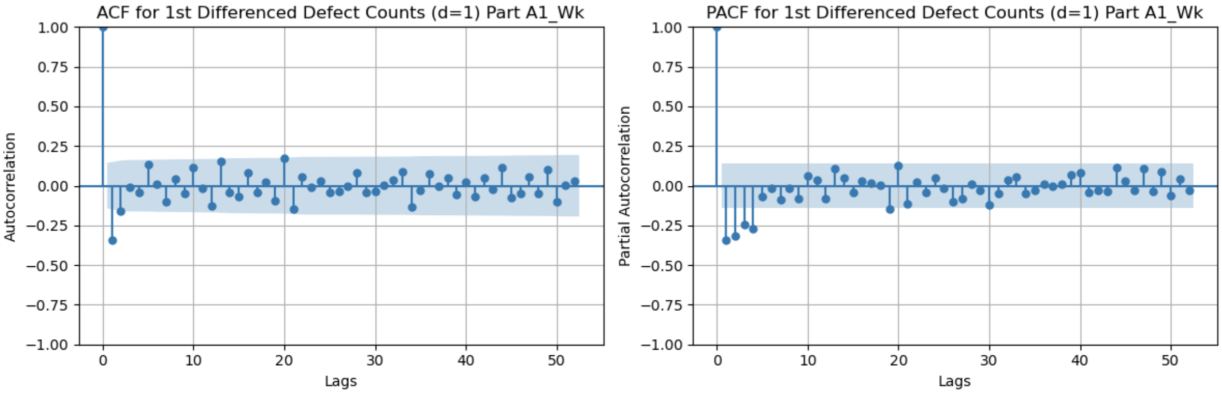


Figure 7.4: ACF and PACF plots for first differenced defect counts (d=1) of Part A1_Wk

Table 7.2: Parameter Ranges for ARIMA Model Grid Search

Part_Freq	p Range	d Range	q Range	Combinations
A1_Wk	0-5	0-1	0-14	180
A1_Mh	0-2	0-1	0-4	30
A2_Wk	0-8	0-1	0-10	198
A2_Mh	0-3	0-1	0-2	24
A3_Wk	0-4	0	0-7	40
A3_Mh	0-3	0-1	0-3	32
B1_Wk	0-14	0	0-14	225
B1_Mh	0-7	0	0-7	64
B2_Wk	0-4	0	0-12	78
B2_Mh	0-3	0-1	0-2	18
C_Wk	0-7	0-1	0-7	96
C_Mh	0-6	0-1	0-6	84

7.2.3. Parameter Optimization

To identify the optimal parameters for the ARIMA model, a comprehensive grid search algorithm is developed for this research, as outlined in algorithm 2 which can be found in Appendix C.2. This method evaluates combinations of p, d, and q parameters within predefined ranges. The process ensures a thorough exploration of the parameter space to capture the best-fitting model.

The grid search algorithm for ARIMA involves the following steps:

1. Manually modifying the ranges for p, d, and q based on the parameters defined in Table 7.2 for specific 'Part_Freq'.
2. Fitting the ARIMA model on the training data for each combination of p, d, and q.
3. Applying a non-negativity constraint to ensure predictions are non-negative.
4. Forecasting defect counts on the test data.
5. Calculating evaluation metrics MSE, MAE, and mMAPE.
6. Normalizing the chosen metrics MSE, MAE, and mMAPE to facilitate comparison.
7. Identifying the best combination of p, d, and q based on an aggregate score of normalized metrics.

This approach ensures that the selected parameters result in the most accurate and reliable forecasts for the defect counts, tailored to the specific characteristics of the data.

7.2.4. Results

The results, summarized in Table 7.3, detail the optimal parameters and performance metrics of the ARIMA models, ranked by normalized MSE, MAE, and mMAPE across twelve different scenarios. The rankings are categorized into groups A, B, and C, separated by dashed lines, with the models within each group ordered by their performance.

Table 7.3: Optimal Parameters and Performance Metrics for ARIMA Model

Part_Freq	p	d	q	Avg. $\hat{y}$	MSE	MAE	mMAPE	Rank
A2_Mh	2	0	0	18.61	3.15	1.29	7.60%	3
A3_Wk	2	0	3	0.49	0.31	0.52	43.30%	7
A2_Wk	5	0	2	3.86	4.26	1.53	34.27%	8
A1_Wk	3	1	4	3.85	5.69	1.72	30.82%	9
A3_Mh	0	1	0	2.00	1.50	0.83	75.00%	11
A1_Mh	0	1	1	16.98	48.66	5.90	34.25%	12
B1_Wk	14	0	12	0.28	0.11	0.28	25.31%	4
B2_Mh	2	1	0	2.27	1.06	0.84	18.63%	5
B2_Wk	1	0	0	0.34	0.55	0.59	35.59%	6
B1_Mh	0	0	5	1.02	0.82	0.64	62.85%	10
C_Mh	1	0	2	0.63	0.06	0.13	4.28%	1
C_Wk	7	1	7	0.11	0.20	0.29	17.02%	2

The ARIMA model for Part C demonstrated superior performance, particularly with monthly data, achieving the lowest MSE, MAE, and mMAPE. This model was ranked first overall, highlighting its robustness despite the limited number of data points. Conversely, Part A1 showed the poorest performance, with high MSE, MAE, and mMAPE values, placing its weekly and monthly models at ninth and twelfth, respectively. The complexity and variability in defect counts for Part A1 posed significant challenges for the ARIMA model. Additionally, the A1 monthly model had a significant outlier at the end of May 2024, where defect counts dropped down from 30 at the end of April to 7 at the end of May, contributing to its high MSE. Figure 7.5 illustrates the performance patterns for Parts A1 and C, with additional plots available in Appendix D: Figure D.4.

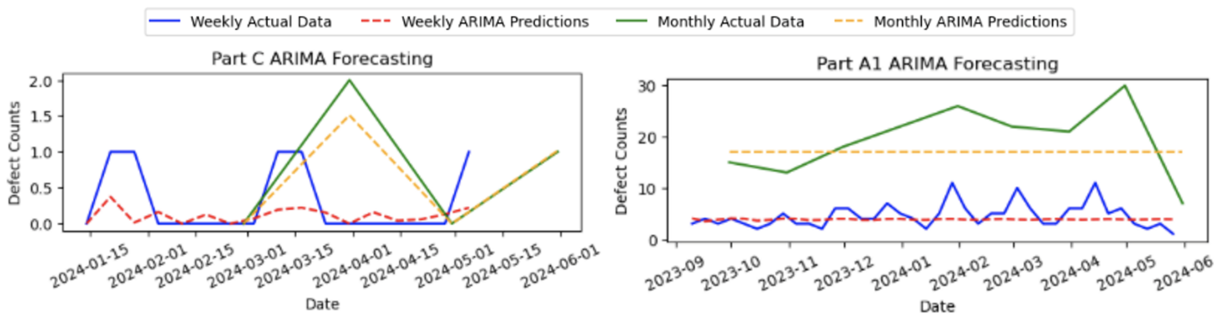


Figure 7.5: Weekly and Monthly ARIMA Results for Part C (rank 1 & 2) and Part A1 (rank 9 & 12)

Monthly data generally performed well, with the ARIMA model for Part C_Mh ranking first and Part A2_Mh third. In the weekly data, Parts C_Wk and B1_Wk ranked second and fourth, respectively. However, these weekly models, despite their high rankings, predicted nearly all values as zero, suggesting that they forecast no defects. While this results in favorable performance metrics, it indicates that the models are primarily predicting non-defects. The same results for both of these parts are similar to the results of the SES model and may not reflect real-world scenarios where some defects are expected, rendering them impractical for forecasting due to the lack of actionable predictions.

The parameters p, d, and q in the ARIMA models offer crucial insights into their performance across different datasets. Models with $d = 0$, such as A2_Mh, B1_Wk, and C_Mh, generally exhibit strong performance, suggesting that these time series are already stationary and lack significant trends. Specifically, B1_Wk, with $p = 14$, $d = 0$, and $q = 12$, ranks highly due to its capability to effectively capture complex autoregressive patterns and noise components.

On the other hand, models with $d = 1$, such as A1_Mh and A3_Mh, reflect the presence of underlying trends necessitating differencing for stationarity. For instance, A1_Wk, characterized by $p = 3$, $d = 1$, and $q = 4$, displays higher complexity because it must account for both trend and noise, leading to moderate performance.

Employing a consistent 80/20 data split across parts provided a balanced training and testing approach. However, the varying number of samples significantly impacts performance metrics. Comparisons between parts with fewer predictions and larger datasets highlight the difficulty in predicting variability, which affects accuracy.

Overall, while the ranking of ARIMA models provides an initial insight into performance, it is challenging to compare metrics across datasets with different characteristics. Therefore, rankings should be interpreted cautiously, with a thorough analysis of what occurs in the forecasts to understand the true performance of the models.

7.3. ARIMAX Model

The ARIMAX model extends the ARIMA model by incorporating external variables that are believed to influence the dependent variable.

7.3.1. Model Configuration

The ARIMAX model can be represented by the equation:

$$y_i = \beta x_i + \sum_{j=1}^{p} \phi_j y_{i-j} + \epsilon_i + \sum_{j=1}^{q} \theta_j \epsilon_{i-j}$$

where y_i is the dependent variable, x_i represents the exogenous variables, ϕ and θ are coefficients for the AR and MA terms, respectively, d indicates the differencing order, and ϵ_i represents white noise. From the library `statsmodels.tsa.statespace.sarimax` the `SARIMAX` function is used to implement the ARIMAX model without seasonality, but with the inclusion of exogenous features as determined in Section 6.2.2.

The ARIMAX model configuration is similar to the ARIMA model but includes exogenous variables to capture additional influences on the time series. These exogenous variables could improve the model's accuracy by accounting for external factors that affect the dependent variable.

7.3.2. Parameter Optimization

Parameter optimization for the ARIMAX model involves running the `SARIMAX` function with the specified exogenous features, which do not require any transformation. It is important to clarify that seasonality is not considered in this context to avoid confusion; there is no standalone ARIMAX model in `statsmodels`, as it is essentially a SARIMAX model where the seasonality parameters are not applied. The optimization process follows similar steps to the grid search algorithm used for the ARIMA model, as defined in Section 7.2.3. Additionally, the same parameter ranges for grid search optimization were employed for the ARIMAX model, as detailed in Table 7.2. The primary distinction lies in the inclusion of "Exogenous Features X_T and X_V" where X_T is utilized for training the model through fitting, and X_V is used in the testing data and predictions. The complete algorithm for the ARIMAX model is outlined in algorithm 3, which can be found in Appendix C.2.

7.3.3. Results

The analysis of the ARIMAX model's performance, as detailed in Table 7.4, reveals the optimal parameters and several key insights. Datasets are ranked by normalized MSE, MAE, and mMAPE across twelve scenarios. Rankings are categorized into groups A, B, and C, separated by dashed lines and ordered by performance within each group.

Models with $d = 0$, such as B1_Wk, C_Mh, and C_Wk, generally perform well, indicating that the time series for these datasets are already stationary and do not exhibit strong trends. For instance, B1_Wk, with parameters $p = 4$, $d = 0$, and $q = 3$, achieves the highest rank with the lowest MSE and MAE, suggesting that the autoregressive and moving average components capture the underlying patterns effectively. Similarly, C_Mh and C_Wk also rank high with $d = 0$, further supporting the absence of a significant trend in these datasets.

Table 7.4: Optimal Parameters and Performance Metrics for ARIMAX Model

Part_Freq	p	d	q	Avg. $\hat{y}$	MSE	MAE	mMAPE	Rank
A3_Wk	3	0	4	0.41	0.30	0.49	38.96%	6
A2_Wk	4	0	1	3.68	3.51	1.40	29.92%	7
A1_Wk	4	0	2	4.38	5.13	1.68	33.81%	8
A2_Mh	0	1	0	17.95	11.67	2.85	15.18%	9
A3_Mh	1	1	3	2.08	1.61	0.91	79.17%	10
A1_Mh	0	1	0	19.32	35.42	4.29	31.07%	12
B1_Wk	4	0	3	0.00	0.09	0.09	4.69%	1
B2_Mh	3	0	0	2.52	0.41	0.48	10.80%	4
B2_Wk	0	0	11	0.40	0.56	0.51	31.38%	5
B1_Mh	1	0	0	1.16	1.13	0.89	85.42%	11
C_Mh	5	0	3	0.68	0.04	0.12	4.31%	2
C_Wk	4	0	6	0.08	0.20	0.26	14.31%	3

Conversely, models with $d = 1$, which are used in A2_Mh, A3_Mh, and A1_Mh, show varying performance. A2_Mh, despite requiring differencing to achieve stationarity, performs moderately well, indicating some underlying trend that the model successfully captures. However, A1_Mh, with the same differencing parameter, performs poorly, suggesting that the trend component in this dataset may not be effectively modeled by the ARIMAX configuration used.

The visual comparison of ARIMAX forecasts with actual defect counts for Part C and Part A1 is shown in Figure 7.6, with additional plots available in Appendix D: Figure D.5. Part C, both weekly and monthly, is well-forecasted, as indicated by the C_Mh line, which closely matches the actual defect count pattern over time. In contrast, the lower-ranked Part A1 exhibits larger deviations, illustrating the challenges in accurately predicting defect counts for this more variable dataset.

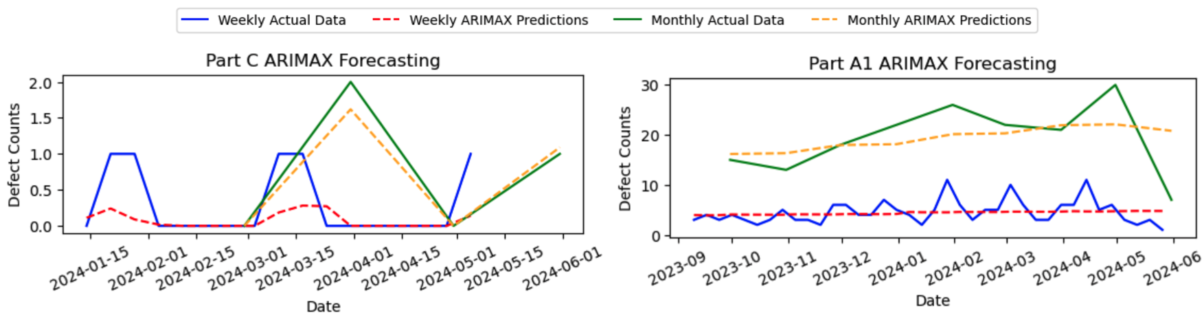


Figure 7.6: Weekly and Monthly ARIMAX Results for Part C (rank 2 & 3) and Part A1 (rank 8 & 12)

The connection between the performance metrics and the exogenous variables, as shown in Table 7.5, provides further insights. The 'Year' variable, where available, shows positive correlations, indicating a trend over time. This is particularly evident in B1_Mh with a correlation of 0.467, where the model also ranks relatively high. In contrast, the 'Month' variable generally exhibits negative correlations, with B2_Mh showing the most extreme value of -1.0, suggesting significant cyclical variations that the model must account for. The strong performance of B2_Mh, with a high negative correlation for 'Month', underscores the importance of capturing these cyclical effects.

The 'Week' variable shows moderate negative correlations across most datasets, reflecting a weekly cyclical pattern in defect counts. The best-performing model, B1_Wk, with a strong negative correlation, highlights the critical role of weekly patterns in forecasting accuracy.

Table 7.5: Pearson Correlation Coefficients for Exogenous Variables (ARIMAX)

Exogenous Variable	A1_Wk	A2_Wk	A3_Wk	B1_Wk	B2_Wk	C_Wk
Year	0.239	NaN	0.039	0.246	NaN	NaN
Month	-0.239	-0.193	-0.071	-0.252	-0.008	-0.091
Week	-0.234	-0.185	-0.058	-0.261	-0.055	-0.105
Holiday_in_Week	-0.096	-0.530	0.113	0.360	-0.255	-0.299
Exogenous Variable	**A1_Mh**	**A2_Mh**	**A3_Mh**	**B1_Mh**	**B2_Mh**	**C_Mh**
Year	0.318	NaN	0.131	0.467	NaN	NaN
Month	-0.373	-0.717	-0.192	-0.509	-1.000	0.135
Holiday_in_Month	0.465	0.345	-0.263	-0.067	0.866	0.302

Holiday-related variables exhibit mixed impacts on defect counts. For instance, A2_Wk shows a strong negative correlation for 'Holiday_in_Week', indicating a significant drop in defect counts during holiday weeks. Conversely, B1_Wk shows a positive correlation for the same variable, suggesting an increase in defects during holiday periods. This variability emphasizes the context-specific influence of holidays on defect counts, necessitating tailored model configurations for different datasets.

Overall, the Pearson correlation coefficients reveal the significance of different exogenous variables in predicting defect counts. The varying correlations highlight the need for customized ARIMAX models that consider specific temporal patterns and external influences. These insights are essential for refining ARIMAX models and enhancing their predictive accuracy in forecasting return flows of defective parts in the semiconductor industry.

7.4. LSTM Model

Long Short-Term Memory (LSTM) model, a specialized form of Recurrent Neural Networks (RNNs), excel in time series prediction by effectively modeling long-term dependencies within sequential datasets. This research utilizes an LSTM network to predict the return flow of defective parts. The model integrates historical defect data along with the relevant exogenous variables, offering a sophisticated forecasting method.

7.4.1. Model Configuration

In configuring the LSTM model for forecasting defect counts of different parts, `TensorFlow` and `Keras` are selected for their robust capabilities. `TensorFlow`, an open-source platform for machine learning, offers powerful tools for building and training models, making it an excellent choice for research experiments. Its extensive library and support for distributed computing allow for efficient handling of large datasets and complex computations, which are often required in research settings. `Keras`, integrated within `TensorFlow`, simplifies the creation of deep learning models, making it ideal for complex forecasting tasks. Its straightforward interface accelerates the model development process, enabling rapid experimentation and iteration, which is crucial for research environments where testing multiple hypotheses and models is common.

The LSTM network for this experiments is built using `Keras` and the Sequential model. This setup allows for the addition of multiple LSTM layers sequentially, followed by a dense (fully connected) layer to produce the final forecast. This configuration is essential for understanding the complex and irregular patterns in defect return data. The Sequential model structure facilitates the stacking of layers, thereby enhancing the model's ability to learn effectively from the data. Its simplicity and flexibility make it a suitable choice for iterative experimentation.

Data pre-processing involved using `MinMaxScaler` from the `Scikit-learn` library to normalize both defect counts and exogenous features, ensuring the data falls within the range of 0 to 1. This normalization accelerates the convergence of the neural network and improves model performance. The exogenous features—year, month, week, and whether there was a holiday in the week or month—provide additional context that could influence defect occurrences. By scaling these features separately, the pre-processing ensured that each type of data was appropriately handled, enhancing the model's ability to learn effectively.

Unlike traditional statistical models, machine learning models require a validation split in addition to the training and test data splits. To maintain comparability with other models, a validation set of 10% was used, subtracted from the training data, resulting in a training set that comprises 70% of the total data.

7.4.2. Hyperparameters Range Selection & Optimization

This subsection outlines the methodology used for generating random combinations of parameters to fine-tune LSTM networks. The objective was to explore a diverse range of parameter settings to identify the most effective configurations for the forecasting model, particularly for predicting defect counts in a time-dependent manner.

To effectively tune the LSTM model, several key hyperparameters and the window size parameters were identified, each with a range of potential values informed by prior research, domain-specific knowledge, and trial-and-error investigation on the dataset. The parameters and their respective ranges are summarized in Table 7.6. Additionally, two specific hyperparameters were determined without a range search to reduce computational complexity, as the grid search was already complex to run.

- **Dropout Rate**: Set at 0.1, this regularization technique prevents overfitting by randomly setting a fraction of input units to zero during training. The 0.1 dropout rate strikes a balance between mitigating overfitting and maintaining the network's learning capacity.
- **Dense Layer**: A fully connected (Dense(1)) layer applied after the LSTM layers, typically with a single neuron and a linear activation function. This configuration is essential for producing continuous output values, which aligns with the nature of regression tasks and is crucial for forecasting defective parts.

Table 7.6: Parameter Ranges for Random Search

Parameter	Range
Epochs	40, 80
Window Size	5, 10, 15, 20
Optimizer	Adamax, RMSprop, Nadam
Batch Size	5, 10, 20, 40
LSTM Units	32, 64, 128
LSTM Layers	1, 2, 3
Activation Function	ReLU, Tanh, Swish

The ranges for each parameter were chosen based on their demonstrated effectiveness in previous studies and practical considerations relevant to the forecasting problem. For instance, the number of epochs balances between sufficient training and computational efficiency. The window size values allow for evaluating short-term versus longer-term historical data's impact on prediction accuracy. The optimizers were selected for their adaptive learning rate capabilities, which are crucial for effectively training deep neural networks. RMSprop was chosen for its ability to handle non-stationary targets, making it ideal for time series forecasting. Adam and Nadam were favored for their simplicity and lower computational requirements, offering efficient training alternatives. Batch sizes were varied to assess the trade-off between training stability and speed. LSTM units and layers were chosen to explore different model capacities and depths, and activation functions were tested to determine their impact on learning nonlinear patterns.

To comprehensively explore the parameter space, an initial random grid search was conducted across a broad range of hyperparameters. Given the high dimensionality and potential interactions between parameters, an exhaustive grid search would result in thousands of combinations, making it computationally prohibitive. Therefore, random sampling was employed to balance comprehensive coverage and computational feasibility, increasing the likelihood of identifying optimal or near-optimal settings. The best performing hyperparameters from this random search will be identified and then used to define a more concise range for the optimal grid search in this research.

The random sampling process involved:

- **Random Sampling**: For each parameter, a random value was selected from its predefined range.
- **LSTM Layer Configuration**: Based on the randomly selected number of layers, a corresponding number of LSTM units were randomly assigned. For example, if the number of layers was 2, two random values from the *lstm_units* range were selected and sorted in descending order to form the layer configuration.

This random sampling process was repeated until 60 unique combinations were generated, ensuring uniqueness by cross-referencing with previously generated combinations.

To reduce the time required for the grid search, initial investigations focused on both weekly and monthly frequencies, but only four out of six parts were selected for the initial investigation, excluding Parts A_2 and C_1. This approach led to eight experiments (4 parts × 2 frequencies) to determine the best parameter ranges.

A comprehensive grid search algorithm was developed for this research, as outlined in algorithm 4 in Section C.2. This algorithm involves:

1. Importing pre-generated random combinations of hyperparameters.
2. Creating sequences with exogenous features based on the randomly selected window size.
3. Splitting the data into training, validation, and test sets.
4. Building the LSTM model with the specified number of layers and units, followed by dropout and dense layers.
5. Training the model on the training set and validating it on the validation set.
6. Making predictions on the test set and applying a non-negativity constraint to ensure predictions are non-negative.
7. Calculating evaluation metrics: MSE, MAE, and mMAPE.
8. Normalizing the chosen metrics to facilitate comparison.
9. Identifying the best parameter configuration based on an aggregate score of normalized metrics.

The generated combinations were stored in dataframes for further analysis, ensuring a structured approach for subsequent evaluation and comparison. The outcomes were then analyzed by identifying the top three combinations for all experiments, providing a valuable selection of the best parameters for a more generalized application in the optimal grid search. Table 7.7 lists the best performing hyperparameters ranges, which will be used consistently across all parts and frequencies in the optimal grid search.

For the monthly data, adjustments to window sizes were necessary due to data sparsity resulting from resampling to monthly frequencies. Because of fewer data points, the model could not effectively learn from larger window sizes. These adjustments ensured the model could learn effectively from the aggregated data. A trial-and-error approach identified optimal window sizes for managing sparse data. Table 7.8 summarizes the window sizes used for the monthly data.

Table 7.7: Selected Hyperparameter Ranges

Parameter	Range
Epochs	40, 80
Optimizer	RMSprop
Batch Size	10, 40
LSTM Units	32, 64, 128
LSTM Layers	2, 3
Activation Function	Tanh

Table 7.8: Selected Window Size Ranges

Part_Freq	Window Size Range
All Weekly Parts	5, 10, 15
A1_Mh **&** A3_Mh	5, 10, 15
A2_Mh	5, 7, 9
B1_Mh	2, 3, 4
B2_Mh	1, 2, 3
C_Mh	2, 3, 4

The same steps are used for the optimal grid search, but all combinations are generated first using the hyperparameters from Table 7.7 and Table 7.8. This resulted in 192 different combinations calculated for the six parts at two different frequencies. The only difference in usage of the algorithm 4 is that instead of importing random combinations, the optimal combinations are used in the first step of the algorithm.

7.4.3. Results

The performance of the LSTM model varied significantly across different parts and frequencies, as detailed in Table 7.9. Part C_Wk demonstrated the highest accuracy; however, the average predicted value is nearly zero, suggesting that the model is predicting no defects. While this results in excellent performance metrics, it indicates that the LSTM is forecasting non-defects for this weekly part dataset, which may not reflect real-world scenarios where some defects are expected. Therefore, from a practical perspective, the model for this data frequency and defect pattern may not be as effective despite the seemingly good performance. Conversely, Part C_Mh exhibited lower performance, likely due to the reduced number of data points and larger time intervals in the monthly data, making it challenging for the model to capture short-term variations effectively.

Table 7.9: Optimal Parameters and Performance Metrics for LSTM Model

Part_Freq	Epoch	Window Size	Batch Size	LSTM Units	Avg. $\hat{y}$	MSE	MAE	mMAPE	Rank
A3_Wk	80	5	40	[128, 128, 128]	0.36	0.26	0.36	23.61%	5
A2_Mh	80	7	40	[128, 128, 128]	20.03	9.40	2.03	12.66%	6
A2_Wk	80	15	10	[128, 128, 128]	2.16	4.09	1.41	25.16%	7
A1_Wk	40	5	10	[128, 128]	3.47	5.56	1.69	30.18%	8
A3_Mh	40	15	40	[64, 64, 64]	2.01	0.81	0.83	53.17%	11
A1_Mh	80	5	40	[64, 64, 64]	18.59	44.36	4.81	35.21%	12
B1_Mh	40	15	40	[64, 32]	0.67	0.31	0.37	20.79%	2
B2_Wk	80	10	40	[64, 32]	0.02	0.43	0.39	20.82%	3
B1_Wk	40	10	10	[32, 32, 32]	0.26	0.11	0.27	23.38%	4
B2_Mh	40	3	10	[32, 32]	1.18	3.28	1.65	39.08%	9
C_Wk	80	15	40	[128, 128]	0.01	0.23	0.26	14.28%	1
C_Mh	80	2	40	[128, 64]	1.18	0.65	0.73	52.11%	10

Part A1, with its higher variability and frequency of defects, posed significant challenges for the LSTM model. Both weekly and monthly forecasts for Part A1 ranked lower, indicating difficulties in capturing the complex and volatile patterns of defect occurrences. This suggests that standard LSTM configurations may struggle with highly variable data and might benefit from additional feature engineering or alternative modeling approaches. Part B1 showed good performance for both weekly and monthly frequencies, reflecting the model's ability to handle datasets with more consistent patterns. Similarly, Part B2_Wk exhibited robust performance, demonstrating the model's effectiveness in capturing defect patterns with an appropriate configuration.

The visual comparison of LSTM forecasts with actual defect counts for Part C and Part A1 reinforces these findings and is shown in Figure 7.7, with additional plots available in Appendix D: Figure D.6. The C_Wk line illustrates that the model forecasts constant zero values. In the right graph, both the weekly and monthly predictions for Part A1 exhibit significant deviations from the actual defect counts, highlighting the challenges in forecasting highly variable and complex datasets.

The Pearson correlation coefficients, shown in Table 7.10, reveal the significance of different exogenous variables in predicting defect counts. The connection between performance metrics and Pearson correlations highlights the influence of exogenous variables on model outcomes. High positive correlations with the 'Year' variable suggest a trend that the LSTM model captures well, particularly for datasets like A1_Wk and A1_Mh. Conversely, negative correlations with 'Month' and 'Week' variables indicate significant cyclical patterns, which the model must account for to improve accuracy.

7.5. Computational Effort of Models

The computational effort required for parameter optimization of various forecasting models is significant, especially given the trial-and-error nature of grid searches. Table 7.11 provides a concise overview of the time taken to run grid searches for each model, emphasizing computational efficiency. Note that

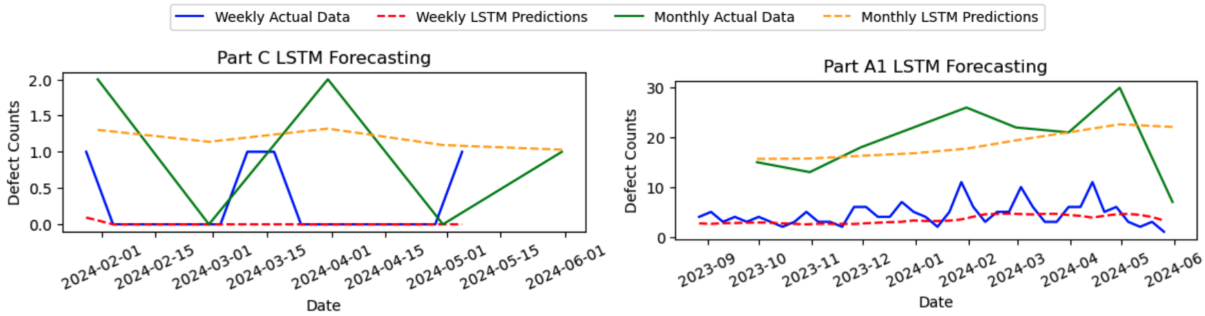


Figure 7.7: Weekly and Monthly LSTM Results for Part C (rank 1 & 10) and Part A1 (rank 8 & 12)

Table 7.10: Pearson Correlation Coefficients for Exogenous Variables (LSTM)

Exogenous Variable	**A1_Wk**	**A2_Wk**	**A3_Wk**	**B1_Wk**	**B2_Wk**	**C_Wk**
Year	0.860	NaN	0.581	-0.511	NaN	NaN
Month	-0.760	-0.406	-0.584	0.535	0.927	-0.518
Week	-0.759	-0.432	-0.575	0.544	0.969	-0.433
Holiday_in_Week	0.246	-0.317	0.068	0.324	0.109	-0.134
Exogenous Variable	**A1_Mh**	**A2_Mh**	**A3_Mh**	**B1_Mh**	**B2_Mh**	**C_Mh**
Year	0.842	NaN	NaN	-0.420	NaN	NaN
Month	-0.635	0.975	-0.902	0.270	-0.452	-0.702
Holiday_in_Month	0.193	-0.219	-0.877	-0.646	-0.054	0.310

the random searches for the LSTM models are not included in this table, but they were already more time-intensive compared to the optimal grid searches of the other models.

Table 7.11: Computational Time of Grid Searches *(in seconds)*

Model	**SES**		**ARIMA**		**ARIMAX**		**LSTM**		**Avg. Part**
Part	*Week*	*Month*	*Week*	*Month*	*Week*	*Month*	*Week*	*Month*	
A1	0.57	9.16	271.08	2.55	100.5	2.35	1,475.17	1,086.26	368.46
A2	0.33	0.18	101.58	1.22	58.06	2.55	1,248.65	930.24	292.10
A3	0.39	0.23	7.88	1.94	6.43	2.07	1,259.99	1,076.36	294.66
B1	0.55	0.24	668.24	13.91	100.07	8.38	2,875.25	2,040.36	713.12
B2	0.25	0.17	51.87	1.46	16.44	2.99	1,831.13	1,550.68	431.12
C	0.29	0.25	35.48	15.06	24.68	10.57	1,564.32	1,477.49	394.52
Avg. Model	0.40	1.70	189.36	6.02	51.03	4.82	1,709.08	1,360.23	416.83

The computational efficiency of forecasting models varies significantly due to their complexity and parameter requirements. SES models are the most efficient, with average grid search times of 0.40 seconds for weekly and 1.70 seconds for monthly data, reflecting their simplicity and limited parameter space. In contrast, ARIMA models, which involve complex time series differencing and parameter estimation, average 189.36 seconds for weekly and 6.02 seconds for monthly data. Adding exogenous variables in ARIMAX models further increases the computational load, averaging 51.03 seconds for weekly and 4.82 seconds for monthly data. LSTM models, with their deep learning architecture, require the most significant computational effort, averaging 1,709.08 seconds for weekly and 1,360.23 seconds for monthly data, due to the extensive data pre-processing, numerous parameters, and multiple epochs involved. These differences underscore the trade-offs between model complexity and computational efficiency, with LSTM models providing the highest accuracy potential at the greatest computational cost.

7.6. Conclusion

This chapter addressed the sixth sub-question of the research:

Research Question 6

What are the optimal parameters for the most suitable forecasting models tailored to different classifications of defective parts and varying data frequencies?

The detailed results for the optimal parameters of each model are presented in their respective sections. Each model's configuration and performance metrics are thoroughly discussed to highlight their effectiveness and limitations for different defect part classifications and data frequencies.

The experiments with SES revealed that while the model is computationally efficient and effective for parts with low variability and infrequent defects, it struggles with parts exhibiting higher variability and more complex patterns. The optimal smoothing constant (α) varied significantly across parts and frequencies, with higher values being more suitable for datasets with more frequent and recent defects.

The ARIMA model, incorporating autoregressive, differencing, and moving average components, demonstrated better performance for datasets requiring differencing to achieve stationarity. However, the complexity and parameter estimation process increased computational times. The optimal parameters for ARIMA varied widely, with higher values of p and q necessary to capture complex and lumpy patterns in some datasets.

The ARIMAX model, which extends ARIMA by incorporating exogenous variables, showed improved performance by accounting for external factors influencing defect counts. The ranges of parameters were similar to the ARIMA model but with differences in optimal parameters due to the inclusion of exogenous variables like 'Year' and 'Month', which enhanced predictive accuracy for some datasets. This model was particularly effective for parts with clear trends and cyclical patterns, as indicated by high Pearson correlation coefficients with the exogenous variables.

The LSTM model provided the most nuanced approach to forecasting defect counts. Its ability to capture long-term dependencies in the data, combined with the inclusion of exogenous features, made it the most effective for parts with complex and volatile patterns. However, this model required significant computational effort. The optimal configurations for LSTM involved multiple layers and units, with the best results obtained using larger batch sizes and longer training epochs.

In conclusion, the optimal parameters for forecasting models varied based on defect part classification and data frequency. SES models were best suited for simple, low-variability datasets, while ARIMA and ARIMAX models handled more complex patterns effectively, especially with the inclusion of exogenous variables. The LSTM model, despite its computational intensity, was most effective for datasets with high variability and long-term dependencies. These findings provide a comprehensive understanding of the strengths and limitations of each forecasting model, guiding the selection of the most appropriate model and parameters for different scenarios in the RSC of defective parts.

8 Models Evaluation

This chapter aims to address the final sub-research question of this thesis:

7. *Which forecasting models provide the highest accuracy in predicting defect counts for specific classifications of defective machinery parts, considering various data frequencies?*

The focus is on evaluating and comparing different forecasting models to determine their accuracy and effectiveness in predicting defects across various part classifications and data frequencies. The detailed analysis includes both disaggregate and aggregate comparisons of model performance, considering weekly and monthly data. Additionally, practical considerations such as computational effort and model complexity are discussed to provide comprehensive insights into the suitability of each model for real-world applications in RSC management within the semiconductor industry.

8.1. Models Comparison

This section presents a detailed comparison of forecasting models for different part classifications, specifically A1, A2, A3, B1, B2, and C. Subsequently, the results are aggregated for classifications A, B, and C. The outcomes for the specific parts on a weekly basis are visualized in Figure D.19 in the Appendix D. Similarly, the graphical performance of the monthly models for these six parts can be compared in Figure D.20.

8.1.1. Disaggregate Comparison of Classified Parts Frequencies

The disaggregate comparison of parts frequencies aims to evaluate the performance of the four forecasting models across different defect part classifications for both weekly and monthly data frequencies. The analysis focuses on determining the most accurate model for each classification and frequency combination, as presented in Table 8.1.

For A1 classification, the LSTM model performed best for weekly data, effectively capturing complex temporal patterns without always predicting zeros. This is logical given the high defect count and extensive dataset, allowing the LSTM to learn both long and short-term dependencies over three years. However, for monthly data, ARIMAX emerged as the best performer, with LSTM showing reduced accuracy due to two main outliers causing high errors. The ARIMA model also performed well, closely matching LSTM's performance for weekly data.

In the A2 classification, ARIMA excelled for monthly data, showing strong performance in capturing trends and patterns, making it ideal for this high defect count dataset. The LSTM model also performed well for weekly data, indicating its efficiency in high-frequency scenarios, although it showed higher errors for monthly predictions.

For the A3 classification, the SES model outperformed other models when evaluated using MSE and MAE metrics for weekly data. However, the SES model consistently predicts almost no defects on a regular basis, which is misleading given the actual defect counts. This inconsistency highlights the limitations of using weekly models for this particular dataset. When examining the monthly data, all models show very high mMAPE values. This is primarily because the models predict a nearly constant defect count of around two each month, which only holds true for two of the four months in question. Therefore, the historical data for this specific A3 part is inadequate for enabling the models to produce accurate predictions.

For B1 classification, ARIMAX and SES were almost equally effective for weekly data, often predicting zero defects due to the dataset's extreme lumpy pattern. The high mean interval between defects led to

Table 8.1: All Models and Part Frequencies Performance Metrics

Part_Freq	Rank	Model	MSE	MAE	mMAPE	Part_Freq	Rank	Model	MSE	MAE	mMAPE
A1_Wk	1	LSTM	5.56	1.69	30.18%	**B1_Wk**	1	ARIMAX	0.09	0.09	4.69%
	2	ARIMA	5.69	1.72	30.82%		2	SES	0.10	0.10	4.69%
	3	SES	5.62	1.71	31.55%		3	LSTM	0.11	0.27	23.38%
	4	ARIMAX	5.13	1.68	33.81%		4	ARIMA	0.11	0.28	25.31%
A1_Mh	5	ARIMAX	35.42	4.29	31.07%	**B1_Mh**	5	LSTM	0.31	0.37	20.79%
	6	LSTM	44.36	4.81	35.21%		6	SES	0.38	0.38	18.75%
	7	SES	46.70	5.74	34.28%		7	ARIMA	0.82	0.64	62.85%
	8	ARIMA	48.66	5.90	34.25%		8	ARIMAX	1.13	0.89	85.42%
A2_Mh	1	ARIMA	3.15	1.29	7.60%	**B2_Mh**	1	ARIMAX	0.41	0.48	10.80%
	3	SES	6.03	1.87	10.89%		3	ARIMA	1.06	0.84	18.63%
	7	LSTM	9.40	2.03	12.66%		7	SES	1.51	0.98	21.33%
	8	ARIMAX	11.67	2.85	15.18%		8	LSTM	3.28	1.65	39.08%
A2_Wk	2	LSTM	4.09	1.41	25.16%	**B2_Wk**	2	LSTM	0.43	0.39	20.82%
	4	ARIMAX	3.51	1.40	29.92%		4	ARIMAX	0.56	0.51	31.38%
	5	SES	3.45	1.50	36.38%		5	ARIMA	0.55	0.59	35.59%
	6	ARIMA	4.26	1.53	34.27%		6	SES	0.61	0.61	35.53%
A3_Wk	1	SES	0.31	0.35	20.95%	**C_Mh**	1	ARIMAX	0.04	0.12	4.31%
	2	LSTM	0.26	0.36	23.61%		2	ARIMA	0.06	0.13	4.28%
	3	ARIMAX	0.30	0.49	38.96%		7	LSTM	0.65	0.73	52.11%
	4	ARIMA	0.31	0.52	43.30%		8	SES	0.71	0.75	47.40%
A3_Mh	5	LSTM	0.81	0.83	53.17%	**C_Wk**	3	ARIMAX	0.20	0.26	14.31%
	6	ARIMA	1.50	0.83	75.00%		4	LSTM	0.23	0.26	14.28%
	7	SES	1.50	0.84	75.00%		5	ARIMA	0.20	0.29	17.02%
	8	ARIMAX	1.61	0.91	79.17%		6	SES	0.30	0.30	14.71%

good scores by predicting no defects, which was not useful. The LSTM model performed best for monthly data, but all models struggled with the dataset's variability.

In the B2 classification, ARIMAX was the most accurate for monthly data, showing a well-predicted pattern and good performance metrics. For weekly data, the LSTM model showed good performance in the MSE and MAE metrics, but its high mMAPE indicates that these results might be misleadingly favorable due to the predominance of non-defective periods. Considering the relatively recent dataset starting just at the end of 2023 and it's defect distribution, ARIMAX emerges as the most reliable choice for monthly predictions of part B2.

For the classification C, the ARIMAX model performed exceptionally well for monthly data (C_Mh). While the ARIMA model also provided very good predictions, ARIMAX had a slightly better overall score. This classification had the smallest dataset, with just 13 defect counts over 563 days and an average interval of 46.9 days between defects. Despite the limited data, the ARIMAX model's predictions were remarkably accurate. The test data included only three defects and two non-defective periods, yet the ARIMAX model, utilizing a complex configuration of AR(5) and MA(3), successfully forecasted these occurrences, demonstrating its robustness even with sparse data.

In conclusion, LSTM models are generally best for high-frequency (weekly) data, especially in classification A, where they capture intricate temporal dynamics effectively. For lower frequency (monthly) data, ARIMA and ARIMAX models excel, providing robust performance in capturing broader trends and cyclical patterns. Classification C highlights ARIMAX's strength in handling sparse data. This nuanced approach ensures better accuracy in predicting defects, enhancing RSC management in the semiconductor industry.

8.1.2. Aggregated Comparison of Classifications

This subsection presents an aggregated comparison of forecasting models for classifications A, B, and C, with the aim of providing a comprehensive evaluation of their performance across both weekly and monthly data. The results for all parts within each classification were averaged to offer an overall assessment of model effectiveness. The aggregated performance metrics for classifications A and B are presented in Table 8.2. Classification C is excluded from this table because it consists of only one part, and its performance metrics are already shown in aggregated form in Table 8.1.

Table 8.2: Average A & B Part Classification All Models and Part Frequencies Performance Metrics

Class_Freq	Rank	Model	MSE	MAE	mMAPE
A_Wk	1	LSTM	3.30	1.15	26.32%
	2	SES	3.13	1.19	29.63%
	3	ARIMAX	2.98	1.19	34.23%
	4	ARIMA	3.42	1.26	36.13%
A_Mh	5	LSTM	18.19	2.56	33.68%
	6	ARIMA	17.77	2.67	38.95%
	7	ARIMAX	16.23	2.68	41.81%
	8	SES	18.08	2.82	40.06%

Class_Freq	Rank	Model	MSE	MAE	mMAPE
B_Wk	1	ARIMAX	0.33	0.30	18.04%
	2	LSTM	0.27	0.33	22.10%
	3	SES	0.36	0.36	20.11%
	4	ARIMA	0.33	0.44	30.45%
B_Mh	5	SES	0.95	0.68	20.04%
	6	ARIMA	0.94	0.74	40.74%
	7	ARIMAX	0.77	0.69	48.11%
	8	LSTM	1.80	1.01	29.94%

In classification A, the LSTM model consistently demonstrated superior performance across both weekly and monthly data. The LSTM model's ability to effectively capture complex temporal patterns is reflected in its lowest MSE and MAE values among the models, indicating its robustness in handling varying temporal resolutions. This makes LSTM particularly well-suited for forecasting in environments where intricate temporal dynamics are crucial.

For classification B, ARIMAX emerged as the most effective model for weekly data, showcasing strong performance in managing high-frequency predictions. However, when applied to monthly data, the SES model outperformed others on average. This performance is misleading, as the SES model often predicted zero defects instead of accurately forecasting future defect occurrences. This tendency to predict zeros, rather than actual defects, leads to seemingly better performance metrics but does not contribute effectively to practical forecasting needs.

While ARIMAX generally demonstrated strong performance, it is noteworthy that the model faced challenges with two specific parts within the classifications, where its accuracy was notably lower compared to the other models. These outliers highlight the model's potential vulnerabilities in certain scenarios. Nevertheless, despite these instances, ARIMAX consistently emerged as the most reliable model for low-frequency (monthly) data, excelling in managing percentage errors and capturing broader trends. This resilience underscores its robustness in most contexts, even when occasional difficulties arise.

A critical factor influencing model performance across all classifications is the variability in defect counts and the occurrence of zero-defect periods. For example, in classifications A3 and B1, where periods with zero defects were frequent, the models often struggled to deliver accurate predictions. This variability complicates the performance metrics and makes direct comparisons across classifications more challenging. The effectiveness of each model is closely tied to the frequency and distribution of defect data available for training and testing, a consideration that must be factored in when interpreting the aggregated results.

The aggregated findings align with the disaggregated analysis, showing that the LSTM model consistently excels in high-frequency (weekly) data, particularly for classification A, where its ability to capture intricate temporal dynamics is especially beneficial. For classification B, ARIMAX performs well in weekly data but shows slightly less accuracy in monthly data. The SES model performs effectively for monthly data in classification B but not in classification A. When considering MAE and MSE, ARIMAX generally proves more effective for lower-frequency (monthly) data due to its capability to capture cyclical patterns. Furthermore, classification C continues to highlight ARIMAX's strength in managing sparse and variable data patterns, as previously discussed in Section 8.1.1.

The overall average performance metrics, as presented in Table 8.3, provide a comprehensive evaluation of each model's strengths and limitations across different data frequencies and classifications. This table offers valuable insights, reinforcing the observations discussed earlier regarding the relative effectiveness of each forecasting model in various contexts.

Table 8.3: Overall Average Performance Metrics for Each Model (Wk and Mh)

Model	Frequency	MSE	MAE	mMAPE
LSTM	Wk	2.6425	1.0275	28.5125%
	Mh	14.625	2.5125	48.0575%
ARIMA	Wk	2.6275	1.03	30.315%
	Mh	14.79	2.465	47.32%
SES	Wk	2.67	1.0625	29.858%
	Mh	14.9875	2.7075	46.98%
ARIMAX	Wk	2.9575	1.1625	29.365%
	Mh	14.03	2.505	43.1575%

However, it is important to consider that the comparison between classifications A and B is influenced by the data split methodology used. Due to the near-annual delivery times of parts, the weekly and monthly models differ significantly in granularity and scope. While the weekly data captures more detailed variations, the monthly data aggregates these into broader trends. This difference in data frequency and the corresponding data splits creates a situation where the models are tested under varying conditions, potentially affecting the comparability of results across classifications. Despite these challenges, the metrics in Table 8.3 provide a valuable overview of model performance, offering important insights for selecting the most appropriate forecasting approach.

8.2. Practical Model Evaluation

This section evaluates the practical applicability of the best-performing forecasting models by considering both their predictive accuracy and computational effort, as discussed in Section 7.5.

From the disaggregate and aggregated comparisons, LSTM models generally excel in high-frequency data, particularly for classifications A and B. For lower frequency data, ARIMA and ARIMAX models are superior, capturing broader trends and cyclical patterns effectively. The computational time for parameter optimization varies significantly among models. SES models are the most computationally efficient, while LSTM models require the most significant computational effort. ARIMA and ARIMAX models fall in between, with ARIMAX being slightly more demanding due to the inclusion of exogenous variables.

Considering both predictive accuracy and computational effort:

- **High-Frequency Data:** LSTM models deliver the highest performance but are computationally intensive. The ARIMA and ARIMAX models may offer a better balance between computational efficiency and accuracy.
- **Low-Frequency Data:** ARIMA and ARIMAX models provide robust accuracy with moderate computational demands. ARIMAX is particularly more effective at capturing complex patterns in sparse data compared to ARIMA.

In conclusion, in resource-constrained environments, ARIMA and ARIMAX models offer a balanced compromise between accuracy and computational efficiency for both frequencies. This approach ensures better defect prediction accuracy, enhancing RSC management in the semiconductor industry.

8.3. Conclusion

This section gives answer to the final sub-research question:

Research Question 7

Which forecasting models provide the highest accuracy in predicting defect counts for specific classifications of defective machinery parts, considering various data frequencies?

For high-frequency (weekly) data, LSTM models are the most accurate, particularly for classification A, due to their ability to capture intricate temporal dynamics effectively. This model's strength lies in handling large datasets with frequent defect occurrences, allowing it to learn both short and long-term dependencies.

For lower-frequency (monthly) data, ARIMA and ARIMAX models perform best, excelling in capturing broader trends and cyclical patterns. These models are particularly well-suited for datasets with infrequent but aggregated defect counts. ARIMAX is especially effective in classifications B and C, where it accurately handles complex and sparse data patterns.

Additionally, the practical evaluation of computational efficiency highlights that while LSTM models are computationally intensive, they offer the highest accuracy for high-frequency data. In contrast, ARIMA and ARIMAX models provide a balanced compromise between accuracy and computational effort for both weekly and monthly frequencies, making them practical alternatives in resource-constrained environments.

9 Conclusion & Discussion

This research demonstrates the effectiveness of advanced time series forecasting models in predicting the return flow volume of defective machinery parts at ASML within the RSC context. This chapter concludes by answering the main research question, comparing findings with existing literature, assessing model complexity and computational demands, and exploring practical applications and limitations.

9.1. Conclusion

By addressing the sub-research questions, the study has identified the most accurate models for different scenarios, leading to the answer to the main research question:

Research Question

How can forecasting models accurately predict the return flow volume of defective machinery parts?

A crucial variable for forecasting is time series-based, determined by the occurrences of defects recorded in the 'Material Notification' (MN). By counting these defects for specific 12NCs based on the timestamps of the MN creation dates, this variable becomes essential for developing a time series forecast model, enabling the prediction of future defects at specific time steps and thereby determining the return flow from the factory in the RSC.

The classification of demand patterns reveals that the majority of parts fall into the "Lumpy" category, indicating irregular and infrequent defect occurrences. This lumpiness and sparseness of data present significant challenges for forecasting models. The ABCD-analysis categorizes parts based on defect frequency and economic value, identifying three parts from classification A, two from classification B, and one from classification C as focal points in this research analysis. This classification highlights the variability in defect occurrences and the economic impact across different parts.

Resampling techniques transform the data into weekly and monthly intervals, reducing noise and enhancing the detection of trends and patterns. This processing step is vital for improving the dataset's suitability for time series forecasting models. For all six selected parts, weekly and monthly data sets are considered, resulting in 12 scenarios to investigate the differences in data variability and forecasting accuracy.

For high-frequency (weekly) data, the LSTM model consistently demonstrates superior accuracy for classification A. This model excels in capturing complex temporal patterns and efficiently handling large datasets characterized by frequent defect occurrences, making it the optimal choice for weekly forecasts. On average, across the three parts analyzed, the LSTM model achieved the best performance for classification A, with an MSE of 3.30, an MAE of 1.15, and an mMAPE of 26.32%.

For classifications B and C, the ARIMAX model consistently outperformed other models for both weekly and monthly data. Its strength lies in its ability to handle complex and sparse data patterns, making it particularly well-suited for these classifications. For B2 monthly data, ARIMAX achieved an MSE of 0.41, MAE of 0.48, and mMAPE of 10.80%. In classification C, the ARIMAX model demonstrated exceptional accuracy, with an MSE of 0.04, MAE of 0.12, and mMAPE of 4.31%. On average, the weekly ARIMAX model also performed better for classification B. However, it can be noted that for both B and C classifications, the models tend to perform more similarly to each other. Overall, when comparing the models across B1, B2, and C parts, the ARIMAX monthly model showed superior performance, resulting in a combined mMAPE of 33.51%, underscoring its effectiveness in these classifications.

In conclusion, the aggregated results are consistent with the disaggregated findings, indicating that LSTM models are generally the best choice for high-frequency (weekly) data, particularly for parts in classification A. This is because LSTM models effectively capture more variable time-related patterns and handle datasets with frequent defect occurrences. For lower frequency (monthly) data, ARIMA and ARIMAX models demonstrate superior performance, as they are adept at capturing broader trends and cyclical patterns. Classification C further highlights ARIMAX's strength in managing sparse data and complex patterns. The evaluation metrics used in this research, MSE, MAE, and mMAPE, demonstrate that both LSTM and ARIMAX models generate relatively small errors and achieve reasonable accuracy in predicting defect counts, particularly for parts with high variability and frequent issues. These predictions can determine the return flow, thereby enhancing ASML's inventory management and capacity planning.

9.2. Discussion

This section provides a interpretation of the results, comparing them with existing literature, and evaluates model complexity and computational demands. Additionally, it explores practical applications, highlights contributions to capacity planning and inventory management, identifies potential areas for improvement, and suggests directions for future research.

9.2.1. Interpretation of Results

The findings underscore the effectiveness of advanced time series forecasting models in predicting the return flow volume of defective machinery parts within ASML's RSC. The LSTM models showed excellent accuracy in predicting high-frequency (weekly) data, particularly for classification A parts. This is because they can effectively capture complex patterns over time and handle large datasets with frequent defect occurrences. For classifications B and C, ARIMAX models were the most effective for both weekly and monthly data, adeptly handling complex and sparse data patterns. To generalize these findings, forecasting defect counts for these six parts should be extended to all parts within the same categories, using the appropriate model for each classification. This approach allows ASML also to determine the total return inflow at any specific time step by summing the predicted defect counts for all parts, thus enhancing inventory and capacity management. However, despite performing excellently in terms of MSE and MAE, some models predicted no defects, which is misleading given the actual defect counts. This highlights the limitations of weekly models for this dataset, suggesting the need for model adjustments or alternative approaches to improve prediction reliability.

9.2.2. Comparison with Existing Literature

This study advances the field of RSC management by addressing significant gaps in the forecasting of return flow volumes, particularly within the semiconductor industry. Previous research, such as Kumar et al. (2014), highlighted the lack of integrative models linking returned product flows with the demand for new parts. This study directly addresses this gap by developing and evaluating specialized forecasting models tailored to ASML's needs.

This study demonstrates that advanced time series forecasting techniques, particularly ARIMAX and LSTM models, are effective for predicting early-stage failures in high-value, complex parts, such as those at ASML. These models effectively manage the sporadic and variable nature of defect occurrences, addressing the limitations of traditional methods like the Weibull distribution, which previous research has found less effective in such cases (Abernethy, 2006). Moreover, the ARIMAX and LSTM models consistently outperformed the SES benchmark model in scenarios where non-zero defect counts were predicted.

Additionally, the introduction of the ABCD-analysis extends the traditional ABC-analysis by categorizing parts based on defect frequency and cumulative counts (Syntetos et al., 2009). This method enhances inventory management by focusing forecasting efforts on the most impactful parts, reducing overstock, and improving overall forecasting accuracy. The study supports the findings of Clottey et al. (2012), showing that ARIMAX outperforms standard ARIMA models when incorporating exogenous variables, further validating its use in complex forecasting scenarios.

However, the study also highlights challenges, particularly the computational demands of LSTM models, as noted by Chandriah and Naraganahalli (2021), and the limitations in generalizability due to the reliance on historical data and a narrow sample of parts. Future research should expand the scope to include more parts and consider additional dynamic variables to further validate and refine these models.

9.2.3. Limitations

This study encountered several limitations that affect the generalizability and practical application of its findings. A key limitation was the reliance on historical defect data, which may not fully capture the range of factors influencing future return flows. Additionally, the computational demands of LSTM models pose challenges for real-time application, requiring significant processing power and time.

The study's focus on high-value, complex parts with a 302-day average planned delivery time provided valuable insights but limited the broader applicability of the findings. This focus hindered the achievement of the initial goal to develop forecasting models for predicting inflows of defective parts and enabling early blocking of new purchases. The models were more effective for capacity planning rather than for reducing new purchases. Additionally, the complexity of accurately blocking new purchases was greater than anticipated, highlighting the need for high precision to ensure sufficient future stock availability. While the ARIMAX model performed well for short-term, monthly predictions aligned with ASML's KPI updates, these forecasts were more suited to operational planning than to achieving the strategic objective of reducing new purchases. This indicates a need for further refinement and expansion of the forecasting models to fully meet the broader goals of the study.

The study also employed an 80/20 data split for training and testing, which, while enhancing model reliability, may not fully address all variables affecting defect occurrences. Additionally, the analysis was limited to six parts, including only one from classification C, which restricts the generalizability of the results. Expanding the analysis to a broader range of parts would provide more comprehensive validation of the models' applicability.

Lastly, certain dynamic variables specific to each equipment number within the cumulative 12NC latest successor were not incorporated, as only historical data was available for them. Exogenous variables such as `Detect Code Description`, `Source of Cause Code Description`, and `Duration of Non-Defect` require future data for effective time series forecasting. Without future values, the models cannot fully function. Although these variables could be explored in future research through regression models, their exclusion limits the scope of this study and suggests a direction for more accurate forecasting in future research.

9.2.4. Implementability

The findings of this research have significant practical implications for ASML's capacity planning and inventory management. The demonstrated robustness of ARIMA and ARIMAX models in generating accurate monthly forecasts makes them particularly suitable for integration into ASML's monthly capacity planning processes. These models align well with the company's operational practices, such as updating KPIs and inventory levels at the beginning of each month, ensuring that inventory is consistently matched with expected return volumes.

While these models are effective for optimizing inventory management, and possibly reducing the risk of excess stock and obsolescence, they are not yet accurate enough for reliably blocking new purchases. Therefore, their implementation should currently focus on enhancing inventory management rather than on proactively preventing new buys. However, the adoption of these forecasting models across a broader range of parts can still be streamlined through systematic and automated processes, minimizing manual intervention. Integrating these models into ASML's existing systems, such as SAP or Spotfire, could significantly improve planning and decision-making processes.

9.2.5. Future Research

Future research should focus on enhancing the robustness and applicability of forecasting models by integrating additional data sources, such as future orders of systems and machine usage patterns. A key area of investigation should be on less complex parts to develop more generalized and accurate forecasts, which can effectively block new purchases and minimize excess inventory. Additionally, it is recommended to explore the application of these models to cumulative defect counts rather than individual parts, which could better support capacity planning by predicting the aggregate return flow from factories or service customers.

Moreover, refining the data split approach beyond the current 80/20 training and testing ratio could capture a broader range of variables affecting defect occurrences. Expanding the analysis to include a more diverse set of parts would enhance the generalizability of the models. Future research should also consider incorporating dynamic exogenous variables, such as `Detect Code Description` and `Source of Cause Code Description`, through regression models or other methods that can account for the lack of future data.

Investigating hybrid models that combine the strengths of LSTM and ARIMAX could lead to improved predictive accuracy. Additionally, applying these models across different areas within the semiconductor industry or similar high-tech sectors would extend their utility. Finally, focusing on scenarios with shorter planned delivery times may allow for better capturing of seasonal variations and trends, ultimately improving the reliability and accuracy of long-term forecasting tools.

10 Recommendations

Recommendation 1: Expand Inventory Coverage
To improve the generalizability and accuracy of the forecasting models, it is essential to include a broader range of parts, particularly those from underrepresented categories such as category C. Expanding the analysis beyond the initial six parts will allow for more comprehensive validation of the models across different defect patterns. This broader coverage will better support inventory management decisions and enhance the reliability of return flow predictions.

Recommendation 2: Focus on Monthly Data Updates
Given the operational focus of ASML's RSC, it is recommended to establish a consistent schedule for updating training data with an emphasis on monthly updates. Regular monthly data refreshes will ensure that the models remain aligned with the latest trends and patterns in defect occurrences, thereby maintaining their accuracy and relevance. This approach supports the operational needs of the RSC Ops department and contributes to more reliable forecasting outcomes.

Recommendation 3: Broaden ABCD Analysis Application
The ABCD analysis method, which has proven effective in prioritizing parts based on defect frequency and economic impact, should be applied across the entire inventory. This broader application will enhance the focus on critical parts, ensuring that resources are allocated efficiently and that forecasting efforts are directed where they are most needed. By prioritizing parts with the highest impact, ASML can optimize inventory management and reduce the risks of stockouts or overstocking.

Recommendation 4: Integrate Exogenous Variables
Future enhancements to the forecasting models should include additional exogenous variables, such as part usage patterns and expected order volumes. Incorporating these variables into a multivariate forecasting framework will provide deeper insights into the factors driving defect occurrences and improve the accuracy of the models. This integration will also help address limitations identified in the current study, leading to more robust and comprehensive forecasting models.

Recommendation 5: Optimize Model Selection and Expand Variable Integration
To enhance both the accuracy and efficiency of defect forecasting, it is recommended to balance the use of LSTM and ARIMAX models based on the available computational resources and data frequency. LSTM models should be employed where high accuracy is needed for high-frequency data, while ARIMAX models should be prioritized in resource-constrained environments due to their efficiency. Additionally, expanding the scope of the models to incorporate a wider variety of parts and dynamic exogenous variables will further improve the predictive power and generalizability of the models. This approach will ensure that the forecasting models remain both practical and robust, enhancing their applicability across different operational scenarios.

Recommendation 6: Aggregate Data for Capacity Planning
To better align the forecasting models with the practical needs of capacity planning, it is recommended to focus on aggregated defect data. Testing models on aggregated defect counts, rather than on individual part data, will leverage larger datasets and improve predictive accuracy for return flow volumes. This approach is particularly relevant for capacity planning, where the total volume of parts is more critical than specific part details.

References

Abernethy, R. (2006). *The New Weibull Handbook*. 5th ed. Florida, United States of America: Abernethy.

Agrawal, S., R. K. Singh, and Q. Murtaza (2015). "A literature review and perspectives in reverse logistics". In: *Resources, Conservation And Recycling* 97, pp. 76–92. DOI: 10.1016/j.resconrec.2015.02.009.

Armstrong, J. S. and F. Collopy (1992). "Error measures for generalizing about forecasting methods: Empirical comparisons". In: *International journal of forecasting* 8.1, pp. 69–80.

Armstrong, J. Scott (2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers.

ASML (2022). *Conflict Materials Report*. https://www.asml.com/en/company/sustainability/responsible-supply-chain. [Accessed 12-03-2024].

ASML Holding N.V. (2015). *Corporate Responsibility Report*.

Bacchetti, Andrea and Nicola Saccani (Dec. 2012). "Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice". In: *Omega* 40, pp. 722–737. DOI: 10.1016/j.omega.2011.06.008.

Benti, Natei Ermais, Mesfin Diro, and Addisu Semie (Mar. 2023). "Forecasting Renewable Energy Generation with Machine Learning and Deep Learning: Current Advances and Future Prospects". In: *Preprints*. Available at Preprints. DOI: 10.20944/preprints202303.0451.v1.

Bowersox, Donald J., David J. Closs, and M. Bixby Cooper (2002). *Supply Chain Logistics Management*. McGraw-Hill.

Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung (2015). *Time series analysis: forecasting and control*. John Wiley and Sons.

Bressanelli, G., M. Perona, and N. Saccani (2019). "Challenges in supply chain redesign for the Circular Economy: a literature review and a multiple case study". In: *International Journal of Production Research* 57.23, pp. 7395–7422. DOI: 10.1080/00207543.2019.1630775.

Brombacher, Ac Aarnout, Pc Peter Sander, Pjm Peter Sonnemans, and Jl Jan Rouvroye (May 2005). "Managing product reliability in business processes 'under pressure'". In: *Reliability engineering and systems safety* 88.2, pp. 137–146. DOI: 10.1016/j.ress.2004.07.003. URL: https://www.sciencedirect.com/science/article/pii/S0951832004001346.

Caniato, F., M. G. M. Kalchschmidt, and S. Ronchi (2011). "Integrating quantitative and qualitative forecasting approaches: organizational learning in an action research case". In: *Journal of the Operational Research Society* 62.3, pp. 413–424. DOI: 10.1057/jors.2010.142.

Cerqueira, Vítor, Luís Torgo, and Carlos Soares (Sept. 2019). *Machine Learning vs Statistical Methods for Time Series Forecasting: Size Matters*. arXiv:1909.13316. URL: https://arxiv.org/abs/1909.13316.

Chai, T. and R. R. Draxler (2014). "Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature". In: *Geoscientific model development* 7.3, pp. 1247–1250.

Chandriah, K. K. and R. V. Naraganahalli (2021). "RNN/LSTM with modified Adam optimizer in deep learning approach for automobile spare parts demand forecasting". In: *Multimedia Tools and Applications* 80.17, pp. 26145–26159.

Chapman, Peter, James Clinton, Randy Kerber, Tom Khabaza, Thomas Reinartz, Colin Shearer, and Rüdiger Wirth (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. Available from SPSS Inc.

Chase, Richard B., F. Robert Jacobs, and Nicholas J. Aquilano (2006). *Operations Management for Competitive Advantage*. McGraw-Hill/Irwin.

Chatfield, C. and M. Yar (1988). "Holt-Winters Forecasting: Some Practical Issues". In: *The Statistician* 37.2, p. 129.

Clottey, T. and W. C. Benton (2010). *Core acquisitions planning in the automotive parts remanufacturing industry*. Tech. rep. Ohio, United States: Ohio State University.

Clottey, T., W. C. Benton Jr, and R. Srivastava (2012). "Forecasting Product Returns for Remanufacturing Operations". In: *Decision Sciences Journal of Innovative Education* 43.4, pp. 589–614.

Cui, H., S. Rajagopalan, and A. R. Ward (2020). "Predicting product return volume using machine learning methods". In: *European Journal of Operational Research* 281.3, pp. 612–627. DOI: 10.1016/j.ejor.2019.05.046.

DeCroix, G. A., J. Song, and P. Zipkin (2009). "Managing an Assemble-to-Order System with Returns". In: *Manufacturing & Service Operations Management* 11.1, pp. 144–159. DOI: 10.1287/msom.1070.0209.

Devarajan, Divya and M. S. Jayamohan (Jan. 2016). "Stock control in a Chemical Firm: Combined FSN and XYZ Analysis". In: *Procedia technology* 24, pp. 562–567. DOI: 10.1016/j.protcy.2016.05.111.

Diener, D. (2004). *Value Recovery from the Reverse Logistics Pipeline*. RAND MG-238-A. Santa Monica, Calif.: RAND Corporation.

Dougherty, M. (1995). "A review of neural networks applied to transport". In: *Transportation Research Part C: Emerging Technologies* 3, pp. 247–260. DOI: 10.1016/0968-090X(95)00009-8.

Driessen, M. M., J. Arts, G. Van Houtum, W. J. W. Rustenburg, and B. Huisman (2014). "Maintenance spare parts planning and control: A framework for control and agenda for future research". In: *Production Planning & Control*, pp. 1–20. DOI: 10.1080/09537287.2014.907586.

Duc, T. T. H., H. T. Luong, and Y. Kim (2010). "Effect of the third-party warehouse on bullwhip effect and inventory cost in supply chains". In: *International Journal of Production Economics* 124.2, pp. 395–407. DOI: 10.1016/j.ijpe.2009.11.036.

Elsaraiti, M. and A. Merabet (2022). "Solar power forecasting using deep learning techniques". In: *IEEE Access* 10, pp. 31692–31698. DOI: 10.1109/ACCESS.2022.3160484.

Fahimnia, B., R. Z. Farahani, and J. Sarkis (2013). "Integrated aggregate supply chain planning using memetic algorithm – A performance analysis case study". In: *International Journal of Production Research* 51.18, pp. 5354–5373. DOI: 10.1080/00207543.2013.774492.

Fleischmann, B., J. van Nunen, M. Speranza, and P. e. Stähly (2012). *Advances in Distribution Logistics*. 1st ed. Vol. 460. London: Springer Science & Business Media.

Fleischmann, M. (2003). "Reverse logistics network structures and design". In: *Business Aspects of Closed-Loop Supply Chains: Exploring the Issues*. Ed. by V.D.R. Jr Guide and L.N. Van Wassenhove. Pittsburgh, PA: Carnegie Mellon University Press.

Gardner, E. S. (1985). "Exponential smoothing: the state of the art". In: *Journal of Forecasting* 4, pp. 1–28.

Geda, M. W. and C. Kwong (2018). "Forecasting of used product returns for remanufacturing". In: *2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. IEEE, pp. 889–893.

Govindan, K., P. Murugesan, Q. Zhu, and D. Kannan (2012). "Analysis of third party reverse logistics provider using interpretive structural modeling". In: *International Journal of Production Economics* 140.1, pp. 204–211. DOI: 10.1016/j.ijpe.2012.01.043.

Groenendijk, J., S. Smits, and D. van Wijngaarden (2020). "The prediction performance of long short-term memory neural networks in a high-frequency setting". In.

Guide, V. D. R. and L. N. Van Wassenhove (2001). "Managing Product Returns for Remanufacturing". In: *Production and Operations Management* 10.2, pp. 142–155. DOI: 10.1111/j.1937-5956.2001.tb00075.x.

— (2006). "Closed□Loop Supply Chains: An Introduction to the Feature Issue (Part 1)". In: *Production and Operations Management* 15.3, pp. 345–350. DOI: 10.1111/j.1937-5956.2006.tb00249.x.

Guide, V. Daniel R. and Luk N. Van Wassenhove (2009). “The evolution of closed-loop supply chain research”. In: *Operations Research* 57.1, pp. 10–18.

Heinecke, Georg, Aris A. Syntetos, and W. Wang (June 2013). “Forecasting-based SKU classification”. In: *International journal of production economics* 143.2, pp. 455–462. DOI: 10.1016/j.ijpe.2011.11.020.

Hochreiter, S. and J. Schmidhuber (1997). “Long short-term memory”. In: *Neural Computation* 9.8, pp. 1735–1780.

Holcomb, D. P. and J. C. North (Jan. 1985). “An infant mortality and long-term failure rate model for electronic equipment”. In: *Nokia Bell Labs Journals and Magazine*. Accessed: date-of-access. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6770821.

Holt, C. C. (1957). “Forecasting seasonals and trends by exponentially weighted moving averages”. In: *International Journal of Forecasting* 5.

Hyndman, R. J. and A. B. Koehler (2006). “Another look at measures of forecast accuracy”. In: *International journal of forecasting* 22.4, pp. 679–688.

Hyndman, Rob J and George Athanasopoulos (2018). *Forecasting: principles and practice*. 2nd ed. https://otexts.com/fpp2/. OTexts. (Visited on 04/14/2024).

Ismail, Aya, Timothy Wood, and Héctor Bravo (Apr. 2018). “Improving Long-Horizon Forecasts with Expectation-Biased LSTM Networks”. In.

Jacobs, O., R. Ichard, and B. Chase (2014). *Operations and Supply Chain Management*. 14th ed. New York: McGraw-Hill Irwin.

Jadhav, Pragati and Maheshwar Jaybhaye (Sept. 2020). “A Manufacturing Industry Case Study: ABC and HML Analysis for Inventory Management”. In: *International journal of research in engineering, science and management* 3.9, pp. 146–149. DOI: 10.47607/ijresm.2020.315.

Jin, Xue-Bo, Jiang Feng Wang, Hui Yan Zhang, and Li Hong Cao (Aug. 2013). “ANFIS Model for Time Series Prediction”. In: *Applied mechanics and materials* 385-386, pp. 1411–1414. DOI: 10.4028/www.scientific.net/amm.385-386.1411. URL: https://doi.org/10.4028/www.scientific.net/amm.385-386.1411.

Johnston, F., J. Boyland, M. Meadows, and E. Shale (1999). “Some Properties of a Simple Moving Average When Applied to Forecasting a Time Series”. In: *Journal of the Operational Research Society* 50.12, pp. 1267–1271.

Kalayci, S. (2003). *SPSS Applied Multivariate Statistical Techniques*. Ankara: Asil Publishing House, pp. 199–259.

Kotler, Philip and Gary Armstrong (2010). *Principles of Marketing*. Pearson Education.

Krapp, M., J. Nebel, and R. Sahamie (2013). “Forecasting product returns in closed-loop supply chains”. In: *International Journal of Physical Distribution & Logistics Management* 43.8, pp. 614–637.

Kumar, D, Hamed Soleimani, and Govindan Kannan (2014). “Forecasting return products in an integrated forward/reverse supply chain utilizing an ANFIS”. In: *International Journal of Applied Mathematics and Computer Science* 24.3, pp. 669–682.

Kumar, S. and V. Putnam (2008). “Cradle to cradle: Reverse logistics strategies and opportunities across three industry sectors”. In: *International Journal of Production Economics* 115.2, pp. 305–315. DOI: 10.1016/j.ijpe.2007.11.015.

Kumar, S. and T. Yamaoka (2007). “System dynamics study of the Japanese automotive industry closed loop supply chain”. In: *Journal of Manufacturing Technology Management* 18.2, pp. 115–138.

Lamghari-Idrissi, D. P. T. (2021). “A New After-Sales Service Measure for Stable Customer Operations”. PhD Thesis 1 (Research TU/e / Graduation TU/e). PhD Thesis. Eindhoven University of Technology.

Lee, Jun-Guel, Tae-Hyeong Kim, Ki Woo Sung, and Sung Won Han (Nov. 2021). “Automobile parts reliability prediction based on claim data: The comparison of predictive effects with deep learning”. In: *Engineering failure analysis* 129, p. 105657. DOI: 10.1016/j.engfailanal.2021.105657.

Lewis, E.E. (1996). *Introduction to Reliability Engineering*. 2nd ed. New York: John Wiley and Sons Inc.

Lu, Ming-Wei (1998). "Automotive reliability prediction based on early field failure warranty data". In: *Quality and Reliability Engineering International* 14.2, pp. 103–108.

Ma, J. and H. M. Kim (2016). "Predictive Model Selection for Forecasting Product Returns". In: *Journal of Mechanical Design* 138.5, p. 054501.

Makridakis, S., E. Spiliotis, and V. Assimakopoulos (2020). "The M5 accuracy competition: Results, findings and conclusions". In: *Int J Forecast*.

Makridakis, Spyros, Steven C. Wheelwright, and Rob J. Hyndman (1998). *Forecasting: Methods and Applications*. John Wiley & Sons.

Marx-Gomez, J., C. Rautenstrauch, A. Nürnberger, and R. Kruse (2002). "Neuro-fuzzy approach to forecast returns of scrapped products to recycling and remanufacturing". In: *Knowledge-Based Systems* 15.1, pp. 119–128.

Matsumoto, Mitsutaka and Shingo Komatsu (2015). "Demand forecasting for production planning in remanufacturing". In: *The International Journal of Advanced Manufacturing Technology* 79, pp. 161–175.

Mishra, A., P. Dutta, S. Jayasankar, P. Jain, and K. Mathiyazhagan (2022). "A review of reverse logistics and closed-loop supply chains in the perspective of circular economy". In: *Benchmarking: An International Journal* 30.3, pp. 975–1020. DOI: 10.1108/BIJ-11-2021-0669.

Molenaers, An, Herman Baets, Liliane Pintelon, and Geert Waeyenbergh (Dec. 2012). "Criticality classification of spare parts: A case study". In: *International journal of production economics* 140.2, pp. 570–578. DOI: 10.1016/j.ijpe.2011.08.013.

Paldino, G. M., J. De Stefani, F. De Caro, and G. Bontempi (2021). "Does AutoML Outperform Naive Forecasting?" In: *Engineering Proceedings*. Vol. 5. Multidisciplinary Digital Publishing Institute, p. 36.

Ravi, V. and R. Shankar (2014). "Reverse logistics: insights from sectoral analysis of Indian manufacturing industries". In: *International Journal of Logistics Systems and Management* 17.2, pp. 234–259.

Roesch, William J. (Dec. 2012). "Using a new bathtub curve to correlate quality and reliability". In: *Microelectronics reliability/Microelectronics and reliability* 52.12, pp. 2864–2869. DOI: 10.1016/j.microrel.2012.08.022.

Rosienkiewicz, M., E. Chlebus, and J. Detyna (2017). "A hybrid spares demand forecasting method dedicated to mining industry". In: *Applied Mathematical Modelling* 49, pp. 87–107. DOI: 10.1016/j.apm.2017.04.027. URL: https://doi.org/10.1016/j.apm.2017.04.027.

Rostami-Tabar, B., M. Z. Babai, A. Syntetos, and Y. Ducq (2013). "Demand forecasting by temporal aggregation". In: *Naval Research Logistics (NRL)* 60.6, pp. 479–498.

Sareminia, S. and F. Amini (2023). "A reliable and ensemble forecasting model for slow-moving and repairable spare parts: Data mining approach". In: *Computers in Industry* 145, p. 103827.

Silver, Edward A, David F Pyke, and Douglas J Thomas (2016). *Inventory and Production Management in Supply Chains*. CRC Press.

Song, H. and G. Li (2008). "Tourism Demand Modelling and Forecasting—A Review of Recent Research". In: *Tourism Management* 29.2, pp. 203–220.

Stoll, J., R. Kopf, J. Schneider, and G. Lanza (2015). "Criticality analysis of spare parts management: a multi-criteria classification regarding a cross-plant central warehousing strategy". In: *Production Engineering*, pp. 225–234.

Suthaharan, Shan (Jan. 2016). *Random Forest learning*, pp. 273–288. DOI: 10.1007/978-1-4899-7641-3\{_}11. URL: https://www.researchgate.net/publication/300333883_Random_Forest_Learning.

Syntetos, A. A., Z. Babai, J. E. Boylan, S. Kolassa, and K. Nikolopoulos (2016). "Supply chain forecasting: Theory, practice, their gap and the future". In: *European Journal of Operational Research* 252.1, pp. 1–26. DOI: 10.1016/j.ejor.2015.11.010.

Syntetos, A. A. and J. E. Boylan (2005). “The Accuracy of Intermittent Demand Estimates”. In: *International Journal of Forecasting* 21.2, pp. 303–314.

Syntetos, Aris A., Kai-Kuang Ma, and Mohamed Zied Babaï (Feb. 2009). “Demand categorisation in a European spare parts logistics network”. In: *International journal of operations and production management* 29.3, pp. 292–316. DOI: 10.1108/01443570910939005.

Temur, G. T. and B. Bolat (2017). “Evaluating efforts to build sustainable WEEE reverse logistics network design: comparison of regulatory and non-regulatory approaches”. In: *International Journal of Sustainable Engineering* 10.6, pp. 358–383.

The Economist (Jan. 2024). *Does Europe at last have an answer to Silicon Valley?* [Accessed: 2024-01-09]. URL: https://www.economist.com/business/2024/01/08/does-europe-at-last-have-an-answer-to-silicon-valley.

Thierry, M., M. Salomon, J. van Nunen, and L.N. van Wassenhove (1995). “Strategic issues in product recovery management”. In: *California Management Review* 37.2, pp. 114–135.

Toktay, Beril, Erwin van der Laan, and Marisa de Brito (May 2003). “Managing Product Returns: The Role of Forecasting”. In: *Working Paper*. DOI: 10.1007/978-3-540-24803-3_3.

Turki, Emna, Oualid Jouini, Ziad Jemai, Laura Urie, Adnane Lazrak, Patrick Valot, and Robert Heidsieck (2022). “Forecasting Extractions in a Closed Loop Supply Chain of Spare Parts: An Industrial Case Study.” In: *International Journal of Production Research*, pp. 236–243.

Van Kampen, T.J., Renzo Akkerman, and Dirk Pieter Van Donk (June 2012). “SKU classification: a literature review and conceptual framework”. In: *International journal of operations and production management* 32.7, pp. 850–876. DOI: 10.1108/01443571211250112.

Wang, C., C. Chien, and A. J. Trappey (2021). “On the Application of ARIMA and LSTM to Predict Order Demand Based on Short Lead Time and On-Time Delivery Requirements”. In: *Processes* 9.7, p. 1157. DOI: 10.3390/pr9071157.

Wang, H., Z. Lei, X. Zhang, B. Zhou, and J. Peng (2019). “A review of deep learning for renewable energy forecasting”. In: *Energy Conversion and Management* 198, p. 111799. DOI: 10.1016/j.enconman.2019.111799.

Winters, P. R. (1960). “Forecasting Sales by Exponentially Weighted Moving Averages”. In: *Management Science* 6.3, pp. 324–342.

Yuan, D. D., M. Li, H. Y. Li, C. J. Lin, and B. X. Ji (2022). “Wind Power Prediction Method: Support Vector Regression Optimized by Improved Jellyfish Search Algorithm”. In: *Energies* 15, p. 6404. DOI: 10.3390/en15176404.

Zeng, J. and W. Qiao (2013). “Short-term solar power prediction using a support vector machine”. In: *Renewable Energy* 52, pp. 118–127. DOI: 10.1016/j.renene.2012.10.009.

A

Scientific Paper

(Scientific paper starts from next page)

Predictive Analytics of Defective Machinery Parts in Reverse Supply Chain: A Case Study at ASML

ing. Bartosz Szarszewski[1]

Abstract
Accurate forecasting in Reverse Supply Chain (RSC) management is crucial for the semiconductor industry, particularly for companies like ASML, which must efficiently manage the return flow of defective machinery parts. This study addresses key gaps by developing and evaluating time series-based forecasting models tailored to ASML's RSC. Using a modified ABC-analysis, parts were categorized based on defect frequency and economic impact, focusing on the most critical components. The research applied and optimized models including SES, ARIMA, ARIMAX, and LSTM, using five years of historical defect data. The analysis showed that LSTM models excel in high-frequency (weekly) forecasts for parts with frequent early-life defects, achieving an average mMAPE of 26.32%. ARIMAX models performed best for lower-frequency (monthly) data, particularly in sparsely represented classifications, with mMAPE as low as 4.31% to 10.80%. Despite a higher mMAPE of 85.42% in one outlier, ARIMAX emerged as the most balanced model, offering a practical trade-off between accuracy and computational efficiency. Furthermore, the study highlights the computational efficiency of ARIMAX, which, although more demanding than SES, provided a favorable balance, with ARIMA and LSTM being more resource-intensive. These findings demonstrate ARIMAX's suitability for long-term forecasting and broader trend analysis, making it the preferred model for ASML's RSC. This research provides a robust, data-driven framework that enhances inventory management and capacity planning, making it possible to predict return flows with low error rates on a monthly basis, particularly in high-tech industries where many defective parts are returned. Future research should incorporate additional dynamic variables, explore hybrid models, and refine data splitting techniques to further improve predictive accuracy and support sustainable supply chain operations.

Keywords
Reverse Supply Chain — Defective Parts — Time Series Forecasting — Return Flow

[1] *MSc Candidate of Transport, Infrastructure, and Logistics - Delft University of Technology, The Netherlands*

1. Introduction

In the high-tech semiconductor industry, RSC management is a critical yet complex aspect of operations, particularly for companies like ASML, a leading original equipment manufacturer (OEM). ASML produces advanced photolithography machines used to manufacture silicon wafers, which are essential components in microchips for major technology firms such as Samsung, TSMC, and Intel. Efficient management of returned defective machinery parts is crucial within this B2B context, where timely and accurate prediction of return flows not only enhances operational efficiency and financial performance but also significantly impacts the circularity of resources within the supply chain, promoting sustainability and reducing waste.

The RSC presents unique challenges that differentiate it from traditional forward supply chains. While forward supply chains primarily rely on demand forecasting driven by market trends and consumer behavior, the RSC focuses on predicting the volume and timing of returns. This objective is complicated by the unpredictable nature of product defects, timing of failures, and customer return behaviors. Research by Kumar et al. emphasizes the critical role of accurately forecasting return volumes in enhancing the profitability and efficiency of RSC [1]. Additionally, their study highlights a significant gap in the literature, noting the absence of integrative research that correlates returned product flows with the demand for new parts. This study aims to address that gap by providing a more comprehensive understanding of these dynamics.

ASML's RSC faces significant challenges due to the high value and complexity of the parts involved, compounded by an overstocked reverse inventory with more than 100,000 returned parts awaiting repair, recycling, or disposal. This issue is particularly acute in the early stages of the parts' lifecycle, where failure rates are undesirably high. Traditional models like the bathtub curve [2], which illustrates high initial failure rates due to "infant mortality", and the Weibull distribution [3], commonly used in reliability engineering, are effective for later-stage failures but fall short in predicting these early failures. Additionally, ASML's RSC currently lacks a forecasting model to provide insights into future inflows of defective parts, leading to excess accumulation, obsolescence, and significant value loss. This absence often results in delays for repairable parts due to the prioritization of new orders, further increasing inefficiencies. Addressing these gaps is essential, particularly due to the high value of defective parts in semiconductor manufacturing and the prevalence of early-stage failures [4].

Given these challenges, there is a need for alternative forecasting models that can deliver more accurate early-stage predictions. Recent studies suggest that models like ARIMA and machine learning techniques provide better accuracy in predicting early-life failures [5]. By leveraging historical data and predictive analytics, this research aims to improve quality control and enhance forecasting accuracy during the early stages of the product lifecycle.

The primary objective of this research is to address the gap in existing literature by developing and evaluating forecasting models specifically tailored to the RSC of the semiconductor industry. This research will focus on predicting the volume of defective machinery parts returned during their early failure stage—a critical yet underexplored area in current forecasting literature. By utilizing ASML's operational data, the study aims to modify and adapt existing predictive models to better suit the unique challenges of the RSC, thereby improving inventory management, reducing overstock, and supporting ASML's sustainability goals. The main research question that this study seeks to answer is:

How can forecasting models accurately predict the return flow volume of defective machinery parts?

The paper is structured as follows: Section 2 outlines the methodology, including system analysis, data retrieval, parts classification, and the selection of forecasting models and evaluation metrics. Section 3 presents the experimental results, focusing on the optimization of parameters for the most suitable forecasting models and their computational efficiency. Section 4 discusses the key findings, and Section 5 concludes with insights and recommendations for future research directions.

2. Methods

This study examines ASML's Closed-Loop Supply Chain (CLSC), which integrates forward and RSC processes to optimize the lifecycle management of machinery parts. The methodology is based on frameworks by [6] and [7], with ASML's CLSC structure illustrated in Figure 1. This diagram shows the flow of parts from OEM suppliers through ASML's factory in Veldhoven, where they are assembled into modules. The research specifically focuses on the area highlighted within the blue rectangle, which represents the initial reverse flow of parts—from the detection of defects at the manufacturing site to the decision-making process regarding repairs, as depicted by the orange lines.

The RSC manages the return flow of defective machinery parts. A key challenge in this process is the unpredictability of return volumes, initiated when a part is flagged as defective through a Defect Notification (DN). This unpredictability complicates inventory management, leading to difficulties in maintaining optimal stock levels and planning for new purchases, potentially resulting in overstocking or stockouts.

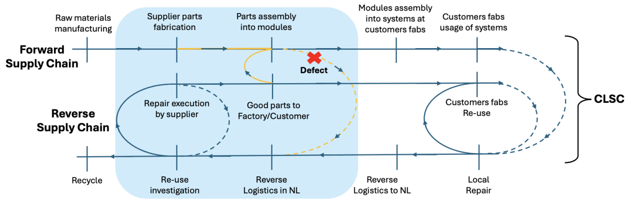


Figure 1. General CLSC of ASML

Further complexities arise from the need to integrate essential data elements, such as Unique Part Codes (UPCs), equipment numbers, and DNs, for effective tracking of defective parts. The variety and complexity of machinery parts across different modules require precise demand triggers and inventory controls.

To address these challenges, the study emphasizes enhancing predictive insights through historical data analysis to identify the key variables and parts to focus on. The most suitable forecasting models are then selected for this case study, followed by the choice of appropriate evaluation metrics. These metrics are used to develop grid searches that determine the optimal parameters for each forecasting model. Ultimately, this approach allows for the evaluation of the most accurate model, providing a comprehensive answer to the main research question. This methodological approach is discussed in the following subsections and systematically displayed Figure 2.

2.1 Data Retrieval

The data is gathered through ASML's internal systems, specifically from the Veldhoven factory, covering the period from May 2019 to May 2024. This five-year dataset then undergoes a process of filtering, cleaning, and analysis to identify the most important variables. The analysis investigates the duration from equipment creation to defect notification, revealing that most defects occur within the first 1000 days of a part's lifecycle, with a significant concentration in the first 100 days. This pattern highlights the prevalence of early-life failures, which is the primary focus of this study. The key variables identified include Unique Part Codes (UPC), which allow for the analysis and forecasting of specific parts' return flows; the Defect Notification (DN), which is the specific code associated with defect occurrences for each part; and the Date of Notification, which serves as the timestamp for the historical DNs. In addition to these, exogenous features such as year, month, week, and public holidays are identified for use in multivariate forecasting models, enabling the capture of cyclical patterns to enhance forecasting accuracy.

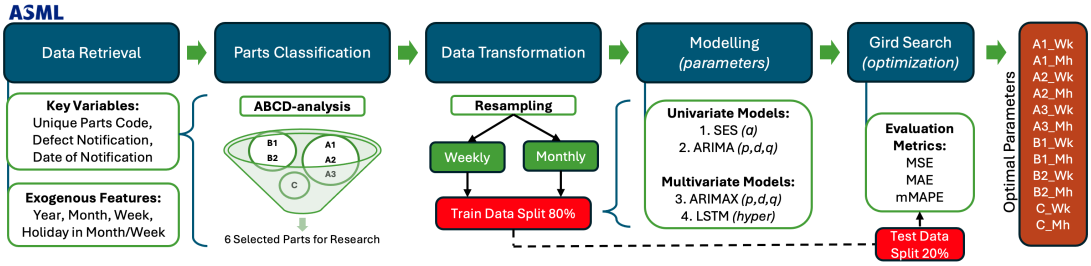


Figure 2. Methodology Flowchart

2.2 Parts Classification

Effective parts classification is crucial for forecasting and inventory management in ASML's RSC. This section integrates insights from literature with a novel methodological approach specifically designed to address the challenges of defect forecasting in semiconductor manufacturing.

Demand Classification: The classification of defective parts follows the framework proposed by [8], which categorizes demand patterns based on Average Demand Interval (ADI) and Coefficient of Variation squared (CV^2). The results indicate that 91.2% of defect notifications fall into the "Lumpy" category, underscoring the sporadic yet highly variable nature of defect occurrences. Additionally, 0.5% of notifications are classified as "Erratic" and 8.3% as "Intermittent", both bordering the threshold for lumpy demand. This predominance of lumpy demand highlights the necessity for specialized classification techniques to accurately forecast return flow volumes.

Inventory Classification: Traditional inventory management methods like ABC-analysis and FSN-analysis are effective for optimizing inventory control by prioritizing parts based on factors such as annual usage value and movement frequency. However, these methods are not directly suited for forecasting defects in semiconductor manufacturing, where defect occurrences across numerous unique parts (UPCs) are irregular and highly variable. Traditional methods typically focus on classifying inventory parts by stock levels and usage rather than addressing the complexity of defect counts.

ABCD-analysis: To address these challenges, this study introduces a novel ABCD-analysis, specifically developed to categorize unique parts (UPCs) based on defect frequency ranges and cumulative defect counts. The ABCD-analysis integrates key elements of ABC-analysis, which classifies parts based on inventory volumes and employs the Pareto distribution to prioritize them by economic value. It extends the traditional ABC framework by adding a fourth Classification "D" to account for parts with low and sporadic defect counts, following the approach of [8] for more nuanced spare parts classification. Additionally, the analysis incorporates aspects of FSN-analysis, where defect occurrences are treated as movements to evaluate frequency. This combined approach ensures a balanced and representative selection of parts for forecasting model application.

The ABCD-analysis reveals that Classification A parts, while representing only 5% of the total unique parts (UPCs), account for 40% of all defects, highlighting their significant impact on return flow and cost management. These parts have defect counts ranging from 39 to 782, as shown in Figure 3, making them a critical focus for this study. Consequently, three parts from Category A (Part A1, Part A2, and Part A3) have been selected for detailed analysis based on their substantial impact and variability in defect counts.

Class.	Defect Count Frequency Range	Percentage of the Unique Parts	Percentage of the Defect Counts
A	[39 – 782]	5%	40%
B	[15 – 38]	14%	30%
C	[5 – 14]	26%	20%
✖ D	[1 – 4]	55%	10%

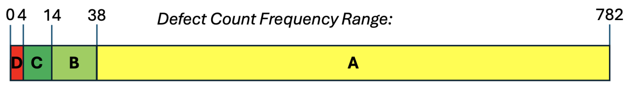


Figure 3. ABCD-analysis

In addition to Classification A, two parts from Classification B (Part B1 and Part B2) and one part from Classification C (Part C) have been chosen for analysis due to their significant, although lower, defect frequencies. Classification B parts exhibit defect counts between 15 and 38, while Classification C parts range between 5 and 14 defects. The exclusion of Classification D is justified by its low impact and the unpredictable nature of its defects, which makes reliable forecasting challenging. Classification D parts, representing 55% of the total but only 10% of the defects, do not provide the consistency needed for effective model training and validation.

The selected parts, detailed in Table 1, are further justified by their high "Impact Scores". The Impact Score is derived by multiplying the economic value of each part by its defect count, where the economic value is calculated using the formula: $Economic\ Value = Demand \times Standard\ Price$. This calculation provides a rational and impactful basis for assessing the economic significance of each part. By combining economic value with defect frequency, the Impact Score offers a comprehensive measure that ensures the most critical parts are prioritized for the analysis. This targeted selection aligns with the research objectives and operational priorities of ASML, enabling a robust and meaningful analysis across different defect frequencies.

Table 1. Statistics for Each Unique Part, Including Defect Counts

Class.	UPC	First DN Date	Last DN Date	Last – First Difference *days*	Mean Duration To Defect *days*	Mean Duration Between DN's *days*	Defect Count
A	Part A1	2020-10-12	2024-05-25	1,321	170.43	3.80	349
A	Part A2	2022-09-30	2024-05-31	609	460.63	2.66	230
A	Part A3	2021-12-06	2024-04-22	868	214.07	15.78	56
B	Part B1	2021-03-09	2024-03-04	1,091	73.72	31.17	36
B	Part B2	2023-03-30	2024-05-30	427	198.50	20.33	22
C	Part C	2022-10-17	2024-05-02	563	160.62	46.92	13

2.3 Data Transformation

Data resampling and splitting are essential for preparing the dataset for time series forecasting. Given the sparse and lumpy nature of the data, aggregating defect occurrences into manageable intervals helps reduce noise and improve forecast accuracy.

Data Resampling: The dataset is resampled on a weekly (Wk) and monthly (Mh) basis to create two time series for each of the six parts under study. Weekly resampling uses the 'W-SUN' option in Python's `Pandas` library, aligning data to each week's end (Sunday). This reduces zero values and smooths daily fluctuations, making it better suited for ASML's RSC operations. Monthly resampling aligns with ASML's practice of updating KPIs and inventory at the start of each month. These resampled datasets form the foundation for the forecasting models, ensuring both practical relevance and an academically sound approach. As shown in the last red block of Figure 2, each unique part (UPC) is renamed according to its frequency (e.g., *A1_Wk* and *A1_Mh*) to standardize references and simplify result interpretation.

Data Split: Following resampling, the data is divided into training (80%) and test (20%) sets, a strategy carefully chosen to balance effective model training with a thorough evaluation of its predictive capabilities. The 80/20 split enables the model to learn from a substantial portion of historical data while reserving a meaningful segment for performance validation on unseen data. This approach not only ensures robust and reliable predictions but also aligns with ASML's operational requirements, facilitating informed decision-making and efficient resource allocation within the RSC.

2.4 Forecasting Models

Selecting the appropriate forecasting models for predicting the return flow of defective machinery parts at ASML required careful consideration of several key requirements, including handling sparse and lumpy data, managing non-stationary trends, and providing accurate medium-term forecasts. The main requirements are summarized in Table 2.

Table 2. Model Selection Requirements

Requirements	
Literature Supported	Time Series Approach
Handling Non-Stationarity	Medium-Term Focus
Complexity vs. Performance	Data Flexibility
Sparse & Lumpy Data	

Model Selection: Based on these requirements, four models were selected: Simple Exponential Smoothing (SES), AutoRegressive Integrated Moving Average (ARIMA), ARIMA model including Exogenous variables (ARIMAX), and Long Short-Term Memory (LSTM) networks.

SES: SES was chosen as a baseline model due to its straightforward implementation and responsiveness to recent data. SES is particularly effective in supply chain environments where up-to-date predictions are critical [9]. Its simplicity makes it a useful benchmark for comparing more complex models.

ARIMA: The ARIMA model was selected for its robustness in handling non-stationary data and its effectiveness in forecasting time series with trends. ARIMA's adaptability makes it well-suited for environments characterized by variability, such as the semiconductor industry. Studies like [10] and [11] demonstrate ARIMA's precision in structured forecasting scenarios. However, ARIMA can struggle with complex dependencies, such as those between past sales and future returns. To address this, the model parameters are carefully optimized, as demonstrated by [5] in their study on forecasting automobile part failures.

ARIMAX: The ARIMAX model extends ARIMA by incorporating exogenous variables, allowing it to account for external factors influencing return flows. This capability is especially valuable in complex manufacturing environments like ASML, where multiple variables impact defect occurrences. The study by [12] highlights ARIMAX's potential to improve forecast accuracy by integrating external influences, making it an ideal choice for this study.

LSTM: The LSTM model is chosen for its ability to manage complex temporal dependencies, particularly in cases involving sparse and irregular data. LSTM is well-suited for handling medium to long-term dependencies, crucial in forecasting defect returns within the semiconductor industry. Research by [5] and [13] underscores LSTM's superior performance in capturing complex patterns compared to traditional models like ARIMA. The study by [5] demonstrated LSTM's effectiveness in predicting automobile part failures using warranty claim data, where its many-to-many approach significantly outperformed other models, highlighting its applicability to industries such as semiconductor manufacturing.

2.5 Evaluation Metrics

Accurately assessing the performance of forecasting models requires selecting appropriate evaluation metrics that can handle the unique challenges of the dataset, such as the presence of zeros and high variability. The selected metrics—Mean Squared Error (MSE), Mean Absolute Error (MAE), and modified Mean Absolute Percentage Error (mMAPE)—are commonly used in similar research and are particularly suited to the characteristics of the data in this study.

2.5.1 Selection of Evaluation Metrics

MSE is particularly useful for highlighting significant forecasting errors due to the squaring of each error term, which makes it sensitive to large deviations. This characteristic is advantageous in scenarios with high variability, as demonstrated in studies by [5] and [14], where it has been effectively applied to models like SES, ARIMA, and LSTM.

MAE is chosen for its simplicity and robustness against outliers. It is indispensable for comparing series measured on the same scale, providing a straightforward calculation and interpretation. MAE ensures that negative and positive errors do not cancel each other out, making it a reliable measure of forecast accuracy, as seen in research involving ARIMA and LSTM models [10, 13].

mMAPE addresses the limitations of standard MAPE, particularly in datasets with zero or near-zero values. By ensuring the denominator is never zero, mMAPE enhances robustness and interpretability, making it suitable for evaluating models in contexts characterized by sparse and lumpy data, such as the ARIMA and LSTM models used in this research [11].

Table 3. Selected Evaluation Metrics

Metric	Equation
MSE	$\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$
MAE	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i-\hat{y}_i\rvert$
mMAPE	$100\% \times \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\lvert\hat{y}_i-y_i\rvert}{1+\lvert y_i\rvert}\right)$

2.5.2 Normalization, Aggregation, and Ranking of Metrics

To ensure fair comparisons of forecasting models using different evaluation metrics, the values of MSE, MAE, and mMAPE are normalized using the min-max normalization technique. This process, similar to the approach used by [5], ensures that each metric is on a comparable scale, facilitating an equitable contribution to the overall performance ranking.

Following normalization, an aggregate score for each model is calculated by averaging its normalized MSE, MAE, and mMAPE scores. Models are then ranked based on their aggregate scores, with the model having the lowest aggregate score considered the best performing. This method allows for a robust comparison of models across different dataset frequencies and part classifications, ensuring the identification of the optimal forecasting model for the research context.

3. Results

3.1 Optimal Parameters for Each Model

To ensure accurate forecasting, the optimal parameters for each selected model— SES, ARIMA, ARIMAX, and LSTM —were determined by developing grid search algorithms. This subsection outlines the key parameters for each model and the approach used to optimize them.

SES primary parameter is the smoothing constant α. A grid search was conducted across α values ranging from 0 to 1 in increments of 0.025. The best α was selected based on the lowest aggregate score derived from normalized evaluation metrics (MSE, MAE, mMAPE).

ARIMA requires selecting the autoregressive order (p), differencing order (d), and moving average order (q). Parameter ranges were determined through ACF, PACF, and ADF tests. A grid search was then performed over p (0-14), d (0-1), and q (0-12) to identify the optimal combination based on the aggregate score of normalized metrics.

ARIMAX extends ARIMA by incorporating exogenous variables, selected through domain knowledge and correlation analysis. The grid search for ARIMAX followed the same procedure as ARIMA, with the inclusion of exogenous variables to improve model accuracy.

LSTM requires optimization of multiple hyperparameters: the number of units, layers, batch size, and window size. After an initial random search, a grid search refined key parameters: epochs (40-80), batch size (10-40), LSTM units (32-128), and layers (1-3). The optimal configurations were chosen based on the lowest aggregate score from normalized metrics.

Table 4. Parameter Ranges for Model Grid Search

Model	Parameter	Range
SES	α	0.0 - 1.0 (increments of 0.025)
ARIMA & ARIMAX	p	0 - 14
	d	0 - 1
	q	0 - 12
LSTM	Epochs	40 - 80
	Window Size	5 - 15
	LSTM Units	32 - 128
	Layers	1 - 3

The grid search process enabled a thorough exploration of the parameter range, resulting in optimal configurations for each forecasting model—SES, ARIMA, ARIMAX, and LSTM—tailored to the dataset's unique characteristics. The selected parameters, evaluation metrics, and rankings for all 12 scenarios, encompassing six unique parts across two frequencies, are thoroughly detailed in the extensive experimental research documentation. These results provide a solid foundation for comparing model performance and ensuring the most accurate forecasting approach for each scenario.

3.2 Models Performance

This section provides a comprehensive evaluation of forecasting models across different classifications (A, B, and C) and data frequencies (weekly and monthly). The objective is to

Table 5. Best Models for Each Classification (Weekly and Monthly)

UPC_Freq	Model	MSE	MAE	mMAPE	UPC_Freq	Model	MSE	MAE	mMAPE
A1_Wk	LSTM	5.56	1.69	30.18%	B1_Wk	ARIMAX	0.09	0.09	4.69%
A1_Mh	ARIMAX	35.42	4.29	31.07%	B1_Mh	SES	0.38	0.38	18.75%
A2_Wk	LSTM	4.09	1.41	25.16%	B2_Wk	LSTM	0.43	0.39	20.82%
A2_Mh	ARIMA	3.15	1.29	7.60%	B2_Mh	ARIMAX	0.41	0.48	10.80%
A3_Wk	SES	0.31	0.35	20.95%	C_Wk	ARIMAX	0.20	0.26	14.31%
A3_Mh	LSTM	0.81	0.83	53.17%	C_Mh	ARIMAX	0.04	0.12	4.31%

identify the most effective models by discussing their performance in specific scenarios, supported by computational efficiency analysis, and to determine the best overall model for defect forecasting in the semiconductor industry.

3.2.1 Model Performance Across Classifications

The performance of the top models for each classification and frequency is summarized in Table 5. This table presents the models that achieved the lowest MSE, MAE, and mMAPE highlighting their effectiveness in different scenarios.

Classification A: The LSTM model consistently demonstrated superior performance for high-frequency (weekly) data, particularly in classification A1, where it effectively captured intricate temporal dynamics. However, for lower-frequency (monthly) data, ARIMAX and ARIMA emerged as the more effective models. For instance, ARIMAX outperformed LSTM in A1_Mh, indicating its strength in capturing broader trends and cyclical patterns.

Classification B: In classification B, ARIMAX was the best performer for weekly data in B1, managing high-frequency predictions with precision. However, for monthly data, the SES model appeared to outperform others in B1_Mh. This result is somewhat misleading, as SES constantly predicted zero defects, leading to artificially favorable performance metrics without practical forecasting value. Despite this, ARIMAX remained more reliable in capturing actual defect occurrences, particularly in B2_Mh, where it provided better MSE and MAE scores.

Classification C: Classification C, characterized by sparse and irregular data, further highlighted ARIMAX's robustness. It consistently outperformed other models in both weekly (C_Wk) and monthly (C_Mh) data, demonstrating its ability to handle limited and irregular data with accuracy.

3.2.2 Computational Efficiency

In addition to accuracy, computational efficiency is a crucial factor in selecting the appropriate forecasting model. Table 6 summarizes the average time required for grid searches across different models, emphasizing the trade-offs between model complexity and computational demands.

The SES model was the most computationally efficient, requiring minimal time due to its simplicity. However, its predictive limitations, particularly the tendency to predict zero defects, make it less suitable for practical forecasting despite

Table 6. Average Computational Time Grid Search

Model	Weekly	Monthly
SES	0.40 sec.	1.70 sec.
ARIMA	189.36 sec.	6.02 sec.
ARIMAX	51.03 sec.	4.82 sec.
LSTM	1,709.08 sec.	1,360.23 sec.

its efficiency. ARIMA and ARIMAX, while more computationally demanding than SES, offered a balanced trade-off between accuracy and efficiency, particularly for monthly data. LSTM, despite its high accuracy in weekly data, was the most computationally intensive, reflecting its deep learning architecture and extensive preprocessing requirements.

3.2.3 Overall Model Evaluation

To provide a comprehensive evaluation, the overall average performance metrics for each model across different frequencies and classifications are presented in Table 7.

Table 7. Overall Average Performance Metrics

Model	Freq.	MSE	MAE	mMAPE
SES	Wk	2.67	1.06	29.86%
	Mh	14.98	2.71	46.98%
ARIMA	Wk	2.63	1.03	30.32%
	Mh	14.79	2.47	47.32%
ARIMAX	Wk	2.96	1.163	29.37%
	Mh	14.03	2.51	43.16%
LSTM	Wk	2.64	1.03	28.51%
	Mh	14.62	2.51	48.06%

The data in Table 7 reinforce that while LSTM models excel in capturing intricate temporal dynamics in high-frequency (weekly) data, their effectiveness diminishes with lower-frequency (monthly) forecasts. In contrast, ARIMA and ARIMAX models consistently perform better for monthly data, with ARIMAX emerging as the most reliable model overall. Its ability to effectively manage broader trends and cyclical patterns, coupled with balanced computational demands, makes ARIMAX the preferred choice for defect forecasting in the semiconductor industry's RSC. This model's robustness in handling sparse and irregular datasets ensures greater accuracy and efficiency, thereby enhancing RSC management.

4. Conclusion

This study sought to address the primary research question: *How can forecasting models accurately predict the return flow volume of defective machinery parts?* By evaluating various forecasting models and introducing a modified ABCD-analysis, the research provides valuable insights for enhancing inventory management and defect forecasting within ASML's RSC.

The analysis highlighted that lumpy demand, characterized by irregular and infrequent defect occurrences, dominates the defect data, underscoring the need for specialized classification techniques to accurately forecast return flows. The ABCD-analysis was introduced as a method to categorize unique parts (UPCs) based on defect frequency and cumulative defect counts, effectively prioritizing parts with the most significant impact on return flow and cost management. This framework provided a solid foundation for applying appropriate forecasting models.

Historical defect occurrences emerged as the key variable in developing accurate time series models. The performance evaluation demonstrated that the LSTM model excelled in high-frequency (weekly) forecasts, particularly for parts in Classification A. These parts, characterized by frequent early-life defects, provided a robust dataset, resulting in an average mMAPE of 26.32%. The inclusion of exogenous variables such as years, months, weeks, and holidays further enhanced the performance of both LSTM and ARIMAX models.

For the more complex and sparsely populated Classifications B and C, which together represent 40% of all parts and 50% of defect counts, the ARIMAX model showed superior performance on a monthly basis. Despite the limited data, 13 defect occurrences for Classification C and 22 for B2, the ARIMAX model achieved a highly accurate mMAPE, ranging from 4.31% to 10.80%, demonstrating its effectiveness in capturing broader trends and handling sparse, irregular datasets. However, the B1 monthly prediction posed challenges, with the ARIMAX model producing an mMAPE of 85.42%. This outlier significantly affected the overall average for Classifications B and C, resulting in a combined mMAPE of 33.51% for the ARIMAX model. Despite this, ARIMAX remained the best-performing model overall for these classifications.

By integrating ABCD-analysis with advanced forecasting models such as ARIMAX and LSTM, ASML can implement a robust strategy for managing its RSC. The ARIMAX model emerges as the most balanced option for practical implementation, offering an optimal trade-off between forecasting accuracy and computational efficiency. While LSTM models are highly effective for high-frequency data, such as weekly forecasts for parts in Classification A, their performance diminishes for lower-frequency forecasts. In contrast, ARIMAX consistently outperforms in scenarios requiring broader trend analysis, making it the preferred choice for defect forecasting in the semiconductor industry's RSC, particularly on a monthly basis. For parts like A1 and A2, where ARIMA and ARIMAX perform comparably well, and for parts in Classifications B and C, a monthly application of the ARIMAX model is recommended. This approach aligns with ASML's practice of updating KPIs and inventory at the start of each month, further enhancing the model's effectiveness.

ARIMAX's ability to deliver accurate predictions, even with limited data, makes it particularly suitable for broader implementation. This model enables ASML and similar industries to optimize inventory management, reduce overstock, and improve supply chain sustainability by accurately forecasting the inflow of returns. Additionally, it could enhance planning for personnel and warehouse capacities, ultimately leading to more informed and strategic decision-making across the organization.

5. Discussion

This study advances the field of RSC management by addressing significant gaps in the forecasting of return flow volumes, particularly within the semiconductor industry. Previous research, such as Kumar et al. [1], highlighted the lack of integrative models that link returned product flows with the demand for new parts, which this study directly addresses by developing and evaluating specialized forecasting models tailored to ASML's needs.

This study demonstrates that advanced time series forecasting techniques, particularly ARIMAX and LSTM models, are effective and usable for predicting early-stage failures in high-value, complex parts, such as those at ASML. These models successfully handle the sporadic and variable nature of defect occurrences, addressing the limitations of traditional methods like the Weibull distribution, which previous research has found less effective in such cases [3]. Furthermore, the ARIMAX and LSTM models outperformed the SES benchmark model in all scenarios where non-zero defect counts were predicted.

Another contribution of this research is the development of the ABCD-analysis, which builds on the traditional ABC-analysis framework by categorizing parts based on defect frequency and cumulative counts [8]. This approach ensures that forecasting efforts are concentrated on the most impactful parts, thus addressing the complexities of inventory management and reducing overstock, especially in industries where part failures are frequent. The combination of ABCD-analysis with advanced forecasting models, particularly LSTM and ARIMAX, provides a robust framework for improving forecasting accuracy, especially for parts with varying defect frequencies.

The LSTM model, known for its ability to manage complex temporal dependencies [5], performed effectively in high-frequency (weekly) forecasts, particularly for Classification A parts where defect occurrences are more frequent. However, the study also highlighted the limitations of LSTM models when applied to lower-frequency data, as these models sometimes predicted no defects despite achieving low MSE and MAE scores, which can be misleading. This finding suggests that while LSTM is powerful in certain scenarios, it may require adjustments or alternative approaches to improve prediction reliability in other contexts.

For the more complex and sparsely populated Classifications B and C, the ARIMAX model, which incorporates exogenous variables such as years, months, weeks, and holidays, demonstrated superior performance. This finding supports the assertion made by [12] that ARIMAX outperforms standard ARIMA models when these additional features are considered, as they significantly enhance forecast accuracy. The impact of these variables varied greatly depending on the specific part and frequency, indicating that a comprehensive approach incorporating all relevant features is necessary for accurate forecasting across all parts in the future. Despite the limited data available for these classifications, the ARIMAX model maintained a high level of accuracy, demonstrating its robustness in capturing broader trends and cyclical patterns. However, while historical defect data played a crucial role, it may not account for all factors influencing return flows, suggesting a limitation in the generalizability of the study's findings.

The study also encountered challenges related to the computational demands of LSTM models, as noted in [13]. These demands limit the feasibility of LSTM for real-time applications due to the significant processing power and time required. Furthermore, the analysis was restricted to only six parts, with limited representation from Classification C, which narrows the scope of the conclusions. This suggests that future research should extend the analysis to a broader range of parts to validate the models more comprehensively.

Additionally, future research should consider incorporating more dynamic exogenous features, such as factory usage patterns and expected orders, into the forecasting models. Expanding the models to a multivariate framework, as recommended by [10], could enhance their predictive capabilities and provide deeper insights into future defect occurrences.

In conclusion, while this study makes significant strides in filling research gaps by developing specialized forecasting models and classification techniques, the identified limitations indicate opportunities for further refinement. Future research should build on these findings by exploring new methodologies, particularly in less complex parts and cumulative defect counts. This approach could enhance the robustness and applicability of forecasting tools, improve capacity planning, and minimize excess inventory within the semiconductor industry. Moreover, investigating hybrid models and refining the data split approach could further advance the accuracy and generalizability of these models across various sectors.

References

[1] D Kumar, Hamed Soleimani, and Govindan Kannan. Forecasting return products in an integrated forward/reverse supply chain utilizing an anfis. *International Journal of Applied Mathematics and Computer Science*, 24(3):669–682, 2014.

[2] E.E. Lewis. *Introduction to Reliability Engineering*. John Wiley and Sons Inc., New York, 2 edition, 1996.

[3] R. Abernethy. *The New Weibull Handbook*. Abernethy, Florida, United States of America, 5 edition, 2006.

[4] William J. Roesch. Using a new bathtub curve to correlate quality and reliability. *Microelectronics reliability/Microelectronics and reliability*, 52(12):2864–2869, 12 2012.

[5] Jun-Guel Lee, Tae-Hyeong Kim, Ki Woo Sung, and Sung Won Han. Automobile parts reliability prediction based on claim data: The comparison of predictive effects with deep learning. *Engineering failure analysis*, 129:105657, 11 2021.

[6] V. D. R. Guide and L. N. Van Wassenhove. Managing product returns for remanufacturing. *Production and Operations Management*, 10(2):142–155, 2001.

[7] M. Thierry, M. Salomon, J. van Nunen, and L.N. van Wassenhove. Strategic issues in product recovery management. *California Management Review*, 37(2):114–135, 1995.

[8] Aris A. Syntetos, Kai-Kuang Ma, and Mohamed Zied Babaï. Demand categorisation in a European spare parts logistics network. *International journal of operations and production management*, 29(3):292–316, 2 2009.

[9] E. S. Gardner. Exponential smoothing: the state of the art. *Journal of Forecasting*, 4:1–28, 1985.

[10] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2 edition, 2018. `https://otexts.com/fpp2/`.

[11] Emna Turki, Oualid Jouini, Ziad Jemai, Laura Urie, Adnane Lazrak, Patrick Valot, and Robert Heidsieck. Forecasting extractions in a closed loop supply chain of spare parts: An industrial case study. *International Journal of Production Research*, pages 236–243, 2022.

[12] T. Clottey, W. C. Benton Jr, and R. Srivastava. Forecasting product returns for remanufacturing operations. *Decision Sciences Journal of Innovative Education*, 43(4):589–614, 2012.

[13] K. K. Chandriah and R. V. Naraganahalli. Rnn/lstm with modified adam optimizer in deep learning approach for automobile spare parts demand forecasting. *Multimedia Tools and Applications*, 80(17):26145–26159, 2021.

[14] R. J. Hyndman and A. B. Koehler. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.

B

Swimlane

(Swimlane figure see next page)

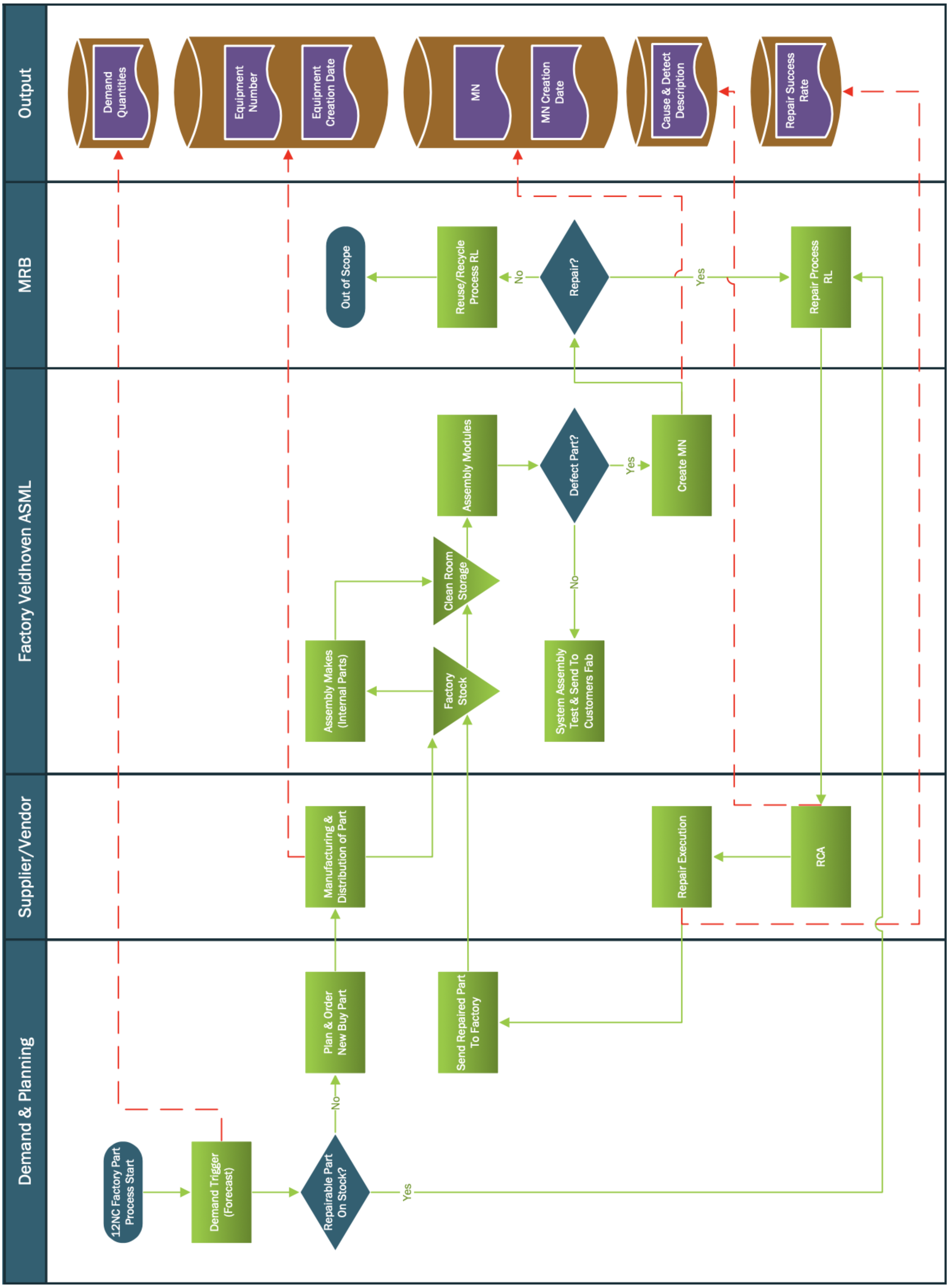


Figure B.1: Swimlane Process of Defective Parts (Authors' own creation)

C

Algorithms

C.1. Grid Search for SES Parameters

Algorithm 1: Grid Search for Exponential Smoothing using `statsmodels`

Init: Training Data T, Test Data V
Hyperparameters: Range for α values to be determined
Result: Best α parameter
for *each* α *in* α_values **do**
 Fit Exponential Smoothing model on training data T *using* α
 $model \leftarrow$ ExponentialSmoothing$(T,$ trend=None, seasonal=None$)$
 Handle possible exceptions
 if *no exception* **then**
  $result \leftarrow model.fit($smoothing_level $= \alpha)$

  Make predictions on test data V
  $predictions \leftarrow result.forecast($steps $=$ len$(V))$
  $predictions \leftarrow$ max(predictions$, 0)$

  Calculate evaluation metrics
  $MSE \leftarrow$ mean_squared_error$(V, predictions)$
  $MAE \leftarrow$ mean_absolute_error$(V, predictions)$
  $mMAPE \leftarrow 100 \times \frac{1}{\text{len}(V)} \sum \frac{|V - predictions|}{1+|V|}$

  Store results for current α
  $results \leftarrow$ append$(results, (\alpha, MSE, MAE, mMAPE))$
 end
end

Normalize and aggregate metrics
for *each metric in* $\{MSE, MAE, mMAPE\}$ **do**
 $normalized\_metric \leftarrow \frac{\text{metric} - \text{min(metric)}}{\text{max(metric)} - \text{min(metric)}}$
 df[Normalized_ $+$ metric] $\leftarrow$ normalized_metric
end

Calculate aggregate score
df[Aggregate_Score] $\leftarrow df$[[Normalized_MSE, Normalized_MAE, Normalized_mMAPE]].mean(axis=1)

Rank models based on aggregate score
df[Rank] $\leftarrow df$[Aggregate_Score].rank()

Find the best parameters
$best\_alpha \leftarrow df$.loc[df[Rank].idxmin()]

return $best\_alpha$

C.2. Grid Search for ARIMA Parameters

Algorithm 2: Grid Search for ARIMA Parameters using `statsmodels`

Init: Training Data T, Test Data V
Hyperparameters: Ranges for p, d, q values to be determined
Result: Best (p, d, q) parameters
for *each* p *in* p_values **do**
  for *each* d *in* d_values **do**
    for *each* q *in* q_values **do**
      Fit ARIMA model on training data T *using* (p, d, q)
      $model \leftarrow \text{ARIMA}(T, \text{order} = (p, d, q))$
      Handle possible exceptions
      if *no exception* **then**
        $result \leftarrow model.fit()$

        Make predictions on test data V
        $predictions \leftarrow result.forecast(\text{steps} = \text{len}(V))$
        $predictions \leftarrow \max(\text{predictions}, 0)$

        Calculate evaluation metrics
        $MSE \leftarrow \text{mean\_squared\_error}(V, predictions)$
        $MAE \leftarrow \text{mean\_absolute\_error}(V, predictions)$
        $mMAPE \leftarrow 100 \times \frac{1}{\text{len}(V)} \sum \frac{|V - predictions|}{1+|V|}$

        Store results for current (p, d, q)
        $results \leftarrow \text{append}(results, (p, d, q, MSE, MAE, mMAPE))$
      end
    end
  end
end

Normalize and aggregate metrics
for *each metric in* $\{MSE, MAE, mMAPE\}$ **do**
  $normalized\_metric \leftarrow \frac{\text{metric} - \min(\text{metric})}{\max(\text{metric}) - \min(\text{metric})}$
  $df[\text{Normalized\_} + \text{metric}] \leftarrow \text{normalized\_metric}$
end

Calculate aggregate score
$df[\text{Aggregate\_Score}] \leftarrow df[[\text{Normalized\_MSE}, \text{Normalized\_MAE}, \text{Normalized\_mMAPE}]]$.mean(axis=1)

Rank models based on aggregate score
$df[\text{Rank}] \leftarrow df[\text{Aggregate\_Score}]$.rank()

Find the best parameters
$best\_params \leftarrow df$.loc[$df$[Rank].idxmin()]

return $best\_params$

C.3. Grid Search for ARIMAX Parameters

Algorithm 3: Grid Search for ARIMAX Parameters using `statsmodels`

Init: Training Data T, Test Data V, Exogenous Features X_T, X_V
Hyperparameters: Ranges for p, d, q values to be determined
Result: Best (p, d, q) parameters
for *each* p *in* p_values **do**
  for *each* d *in* d_values **do**
    for *each* q *in* q_values **do**
      Fit ARIMAX model on training data T *using* (p, d, q) *and exogenous features* X_T
      $model \leftarrow \text{SARIMAX}(T, \text{exog} = X_T, \text{order} = (p, d, q), \text{enforce\_stationarity} = False, \text{enforce\_invertibility} = False)$
      Handle possible exceptions
      if *no exception* **then**
        $result \leftarrow model.fit(disp = False)$

        Make predictions on test data V *using exogenous features* X_V
        $predictions \leftarrow result.forecast(\text{steps} = \text{len}(V), \text{exog} = X_V)$
        $predictions \leftarrow \max(\text{predictions}, 0)$

        Calculate evaluation metrics
        $MSE \leftarrow \text{mean\_squared\_error}(V, predictions)$
        $MAE \leftarrow \text{mean\_absolute\_error}(V, predictions)$
        $mMAPE \leftarrow 100 \times \frac{1}{\text{len}(V)} \sum \frac{|V - predictions|}{1+|V|}$

        Store results for current (p, d, q)
        $results \leftarrow \text{append}(results, (p, d, q, MSE, MAE, mMAPE))$
      end
    end
  end
end

Normalize and aggregate metrics
for *each metric in* $\{MSE, MAE, mMAPE\}$ **do**
  $normalized\_metric \leftarrow \frac{\text{metric} - \min(\text{metric})}{\max(\text{metric}) - \min(\text{metric})}$
  $df[\text{Normalized\_} + \text{metric}] \leftarrow \text{normalized\_metric}$
end

Calculate aggregate score
df[Aggregate_Score] $\leftarrow$ df[[Normalized_MSE, Normalized_MAE, Normalized_mMAPE]].mean(axis=1)

Rank models based on aggregate score
df[Rank] $\leftarrow$ df[Aggregate_Score].rank()

Find the best parameters
$best\_params \leftarrow df$.loc[$df$[Rank].idxmin()]

return $best\_params$

C.4. Random & Grid Search for LSTM Parameters

Algorithm 4: Random and Grid Search for LSTM Parameters using `tensorflow` and `keras`

Init: Training Data T, Validation Data V, Test Data E
Hyperparameters: Ranges for epochs, window size, optimizer, batch size, LSTM units, LSTM layers, activation function
Result: Best parameter configuration
for *each combination of parameters in random/grid search* **do**

Extract parameter values
$epochs, window\_size, optimizer, batch\_size, lstm\_units,$
$lstm\_layers, activation \leftarrow$ parameter combination

Create sequences with exogenous features
$X, y \leftarrow$ create_sequences$(T, exogenous_data, window_size)$

Split data into train, validation, and test sets
$X\_train, y\_train, X\_val, y\_val, X\_test, y\_test \leftarrow$ split_data(X, y)

Build LSTM model
$model \leftarrow$ Sequential()
$model.add($Input$(shape = (window_size, X_train.shape[2])))$
for *each* $units$ *in* $lstm\_units$ **do**
$model.add($LSTM$(units, activation = activation, return_sequences = (i <$ len$(lstm_units) - 1)))$

$model.add($Dropout$(0.1))$
end
$model.add($Dense$(1))$
$model.compile(optimizer = optimizer, loss =' mse')$

Train the model
$model.fit(X\_train, y\_train, epochs = epochs, batch\_size = batch\_size, validation\_data =$
$(X\_val, y\_val), verbose = 0)$

Make predictions for test set
$y\_pred \leftarrow model.predict(X\_test)$
$y\_test\_inv \leftarrow$ scaler_defect.inverse_transform$(y_test.reshape(-1, 1))$
$y\_pred\_inv \leftarrow$ scaler_defect.inverse_transform(y_pred)
$y\_pred\_inv \leftarrow$ np.maximum$(y_pred_inv.flatten(), 0)$

Calculate evaluation metrics
$MSE \leftarrow$ mean_squared_error(y_test_inv, y_pred_inv)
$MAE \leftarrow$ mean_absolute_error(y_test_inv, y_pred_inv)
$mMAPE \leftarrow 100 \times$ np.mean(np.abs$((y_pred_inv - y_test_inv)/(1 +$ np.abs$(y_test_inv))))$

Store results for current parameter combination
$results \leftarrow$ append$(results, (epochs, window_size, optimizer, batch_size, lstm_units,$
$lstm\_layers, activation, MSE, MAE, mMAPE))$
end

Normalize and aggregate metrics
for *each metric in {MSE, MAE, mMAPE}* **do**
$normalized\_metric \leftarrow \frac{\text{metric} - \text{min(metric)}}{\text{max(metric)} - \text{min(metric)}}$
$df[$Normalized_ $+$ metric$] \leftarrow$ normalized_metric
end

Calculate aggregate score
$df[$Aggregate_Score$] \leftarrow df[[$Normalized_MSE, Normalized_MAE, Normalized_mMAPE$]]$.mean(axis=1)

Rank models based on aggregate score
$df[$Rank$] \leftarrow df[$Aggregate_Score$]$.rank()

Find the best parameters
$best\_params \leftarrow df$.loc$[df[$Rank$]$.idxmin()$]$

return $best\_params$

D
Graphs

D.1. Resampled Distribution Graphs

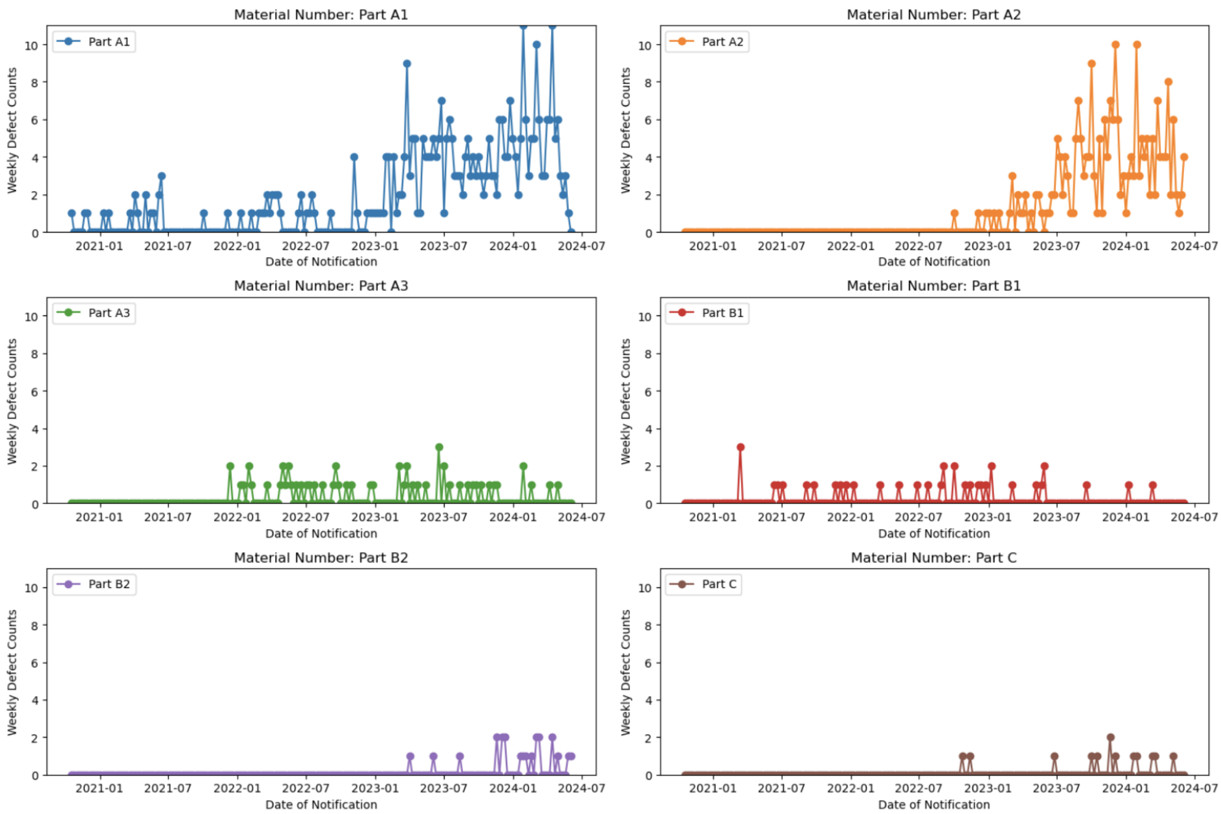


Figure D.1: Weekly Resampled Parts Distribution

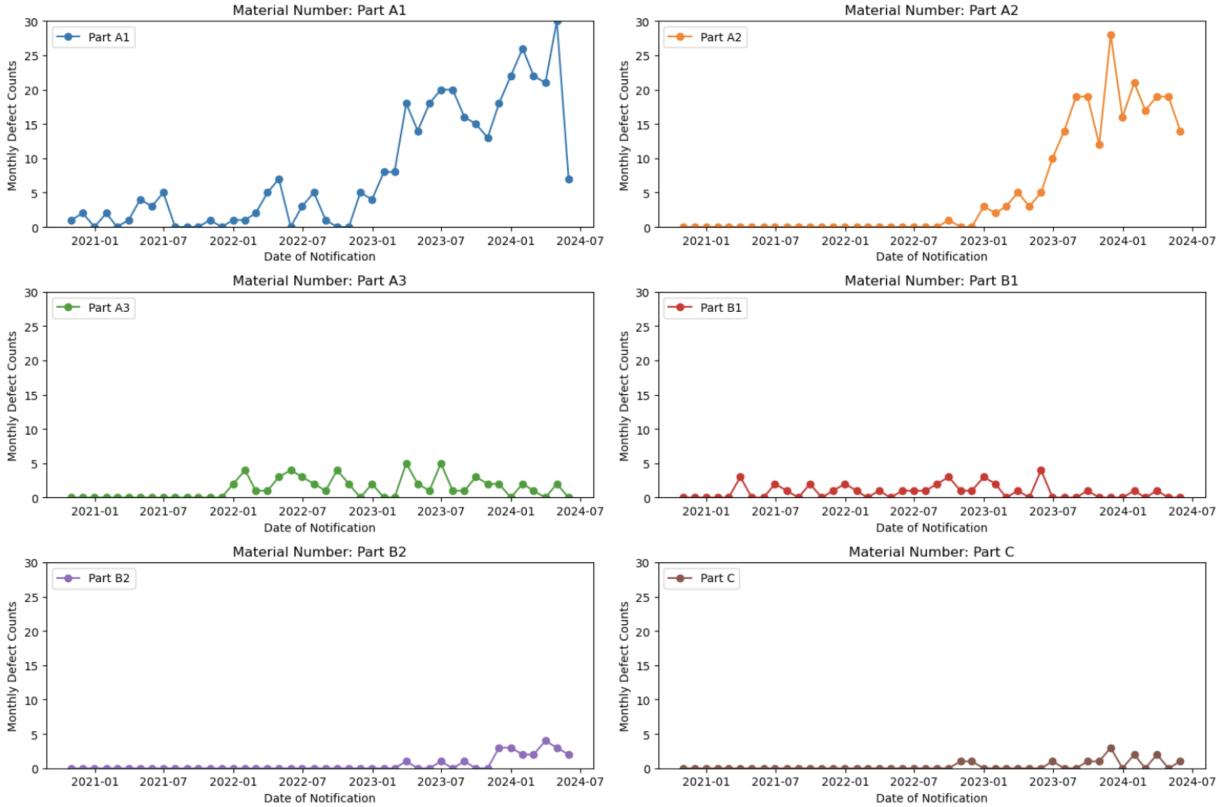


Figure D.2: Monthly Resampled Parts Distribution

D.2. Weekly & Monthly Forecasting Results per Model

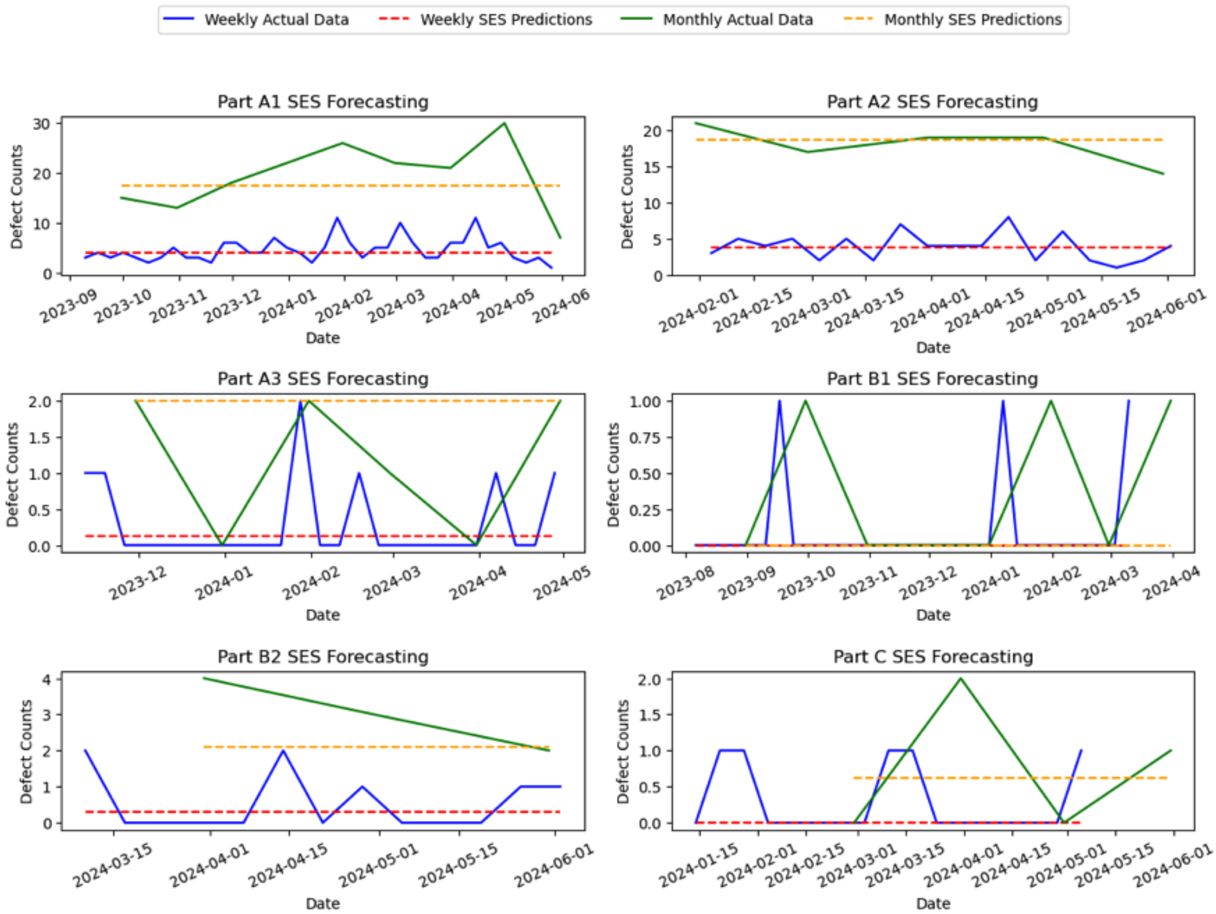


Figure D.3: Weekly and Monthly SES Results

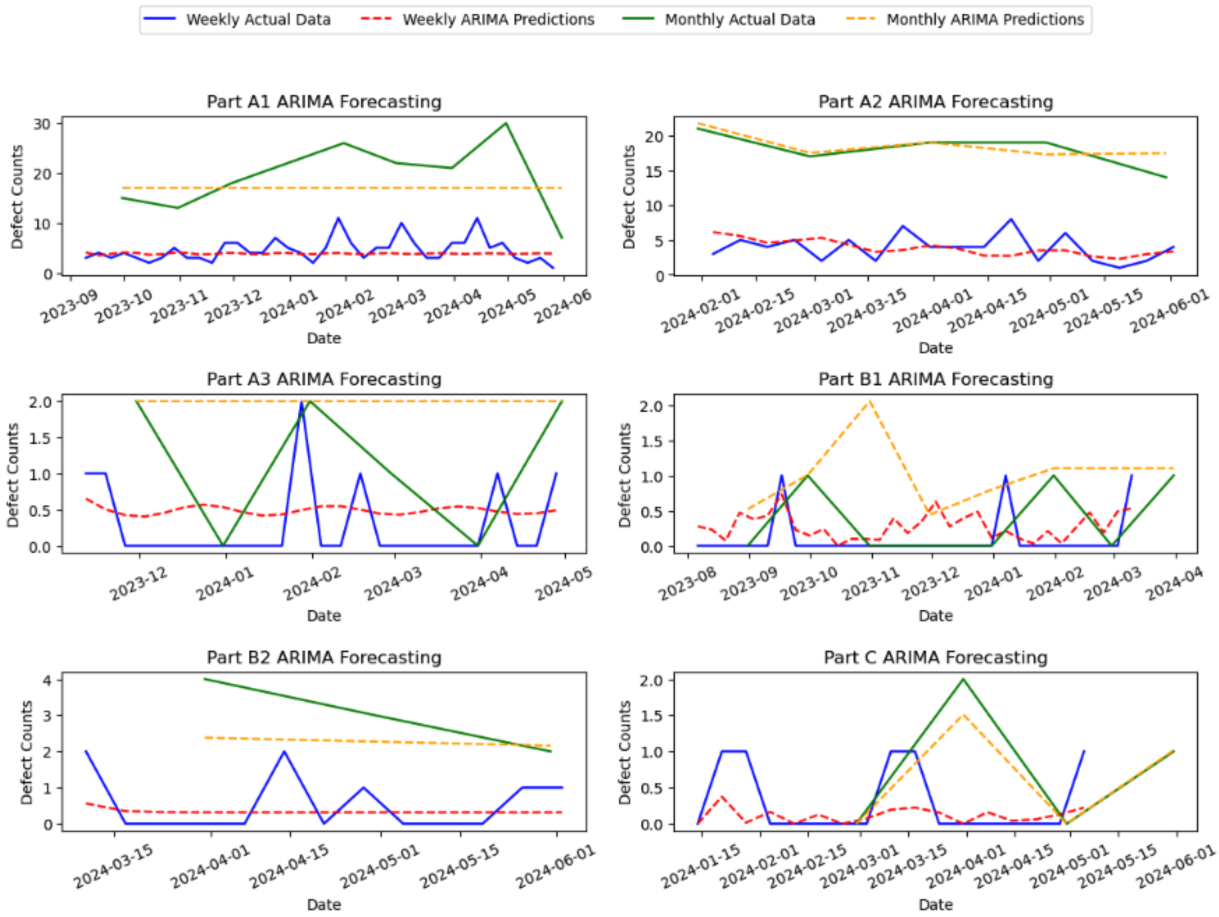


Figure D.4: Weekly and Monthly ARIMA Results

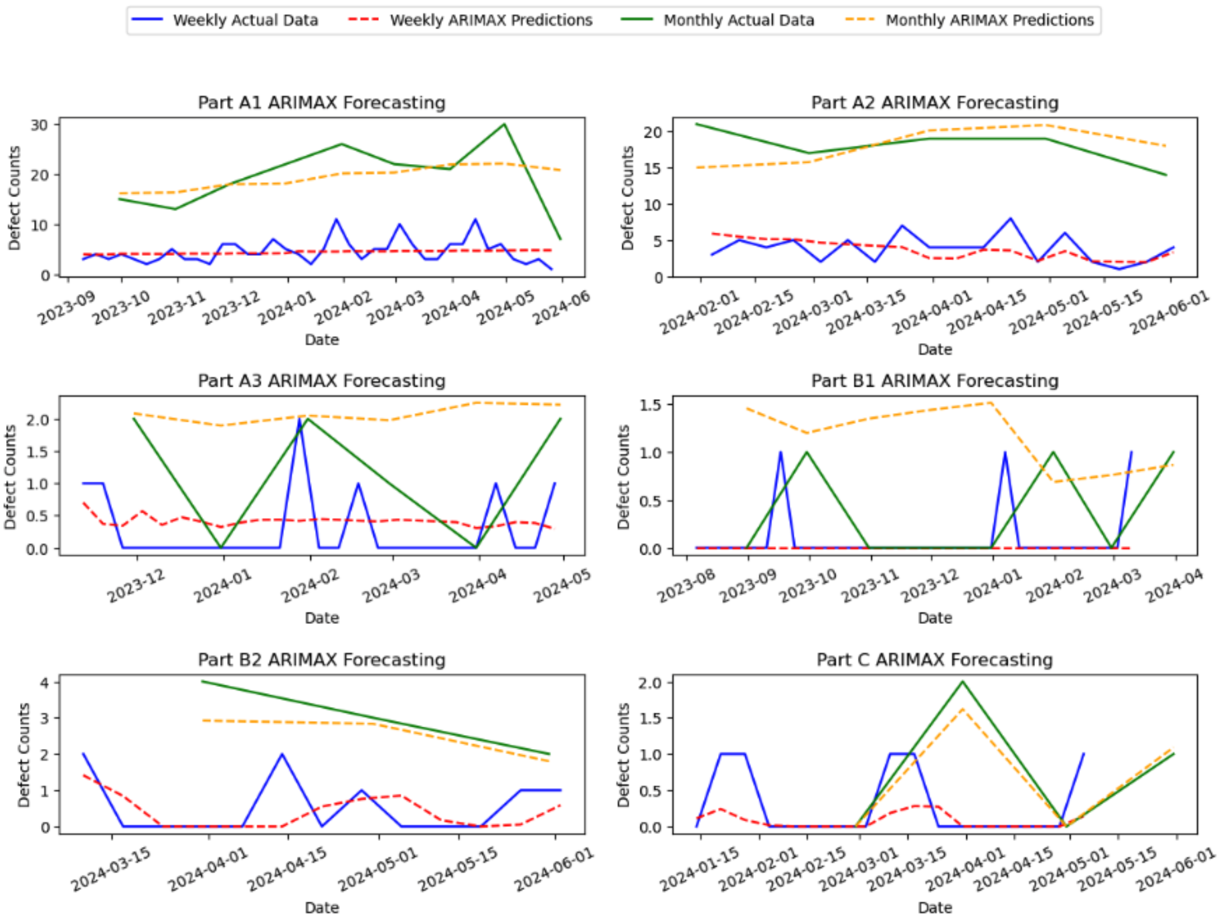
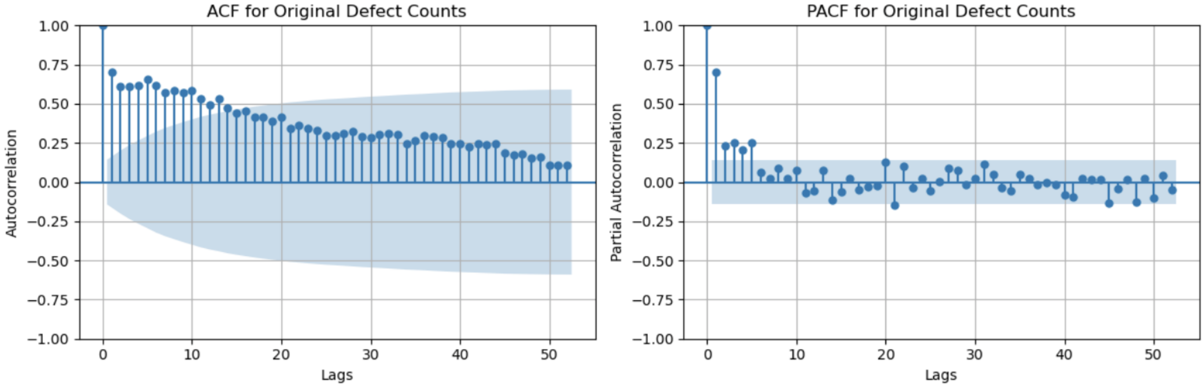


Figure D.5: Weekly and Monthly ARIMAX Results

Figure D.6: Weekly and Monthly LSTM Results

D.3. ACF, PACF, and ADF for each Part_Freq

```
ADF Statistic for Original Series: -1.872838688024978
p-value for Original Series: 0.34494451966646056
The time series Original Series is non-stationary (fail to reject the null hypothesis of the ADF
test).
```




```
ADF Statistic for First Differenced Series: -6.756882131240203
p-value for First Differenced Series: 2.85682926494965e-09
The time series First Differenced Series is stationary (reject the null hypothesis of the ADF tes
t).
```

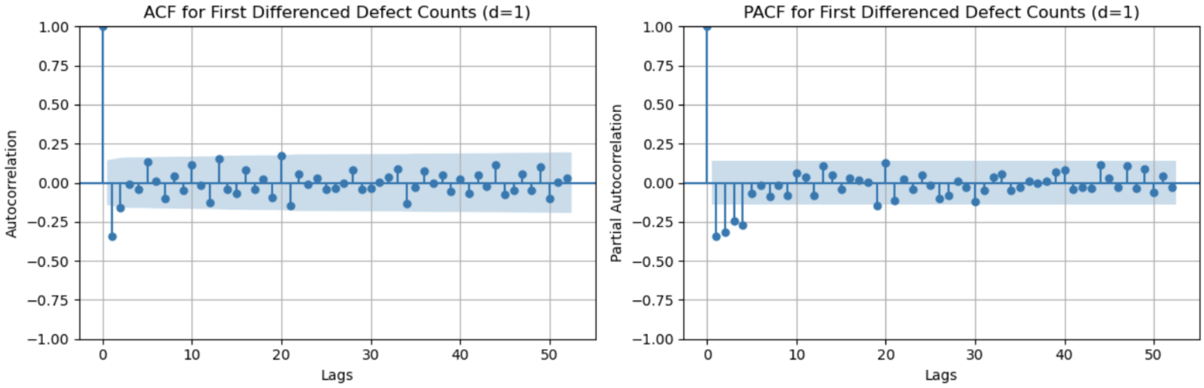
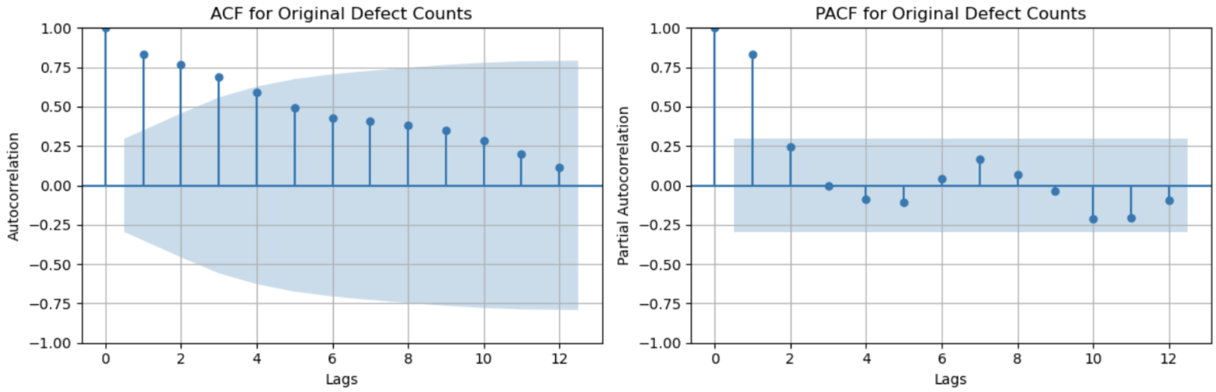


Figure D.7: ACF, PACF, and ADF for Part A1_Wk

```
ADF Statistic for Original Series: -0.9972363259662151
p-value for Original Series: 0.7542709526540534
The time series Original Series is non-stationary (fail to reject the null hypothesis of the ADF
test).
```




```
ADF Statistic for First Differenced Series: -7.85755980966924
p-value for First Differenced Series: 5.391657739082568e-12
The time series First Differenced Series is stationary (reject the null hypothesis of the ADF tes
t).
```

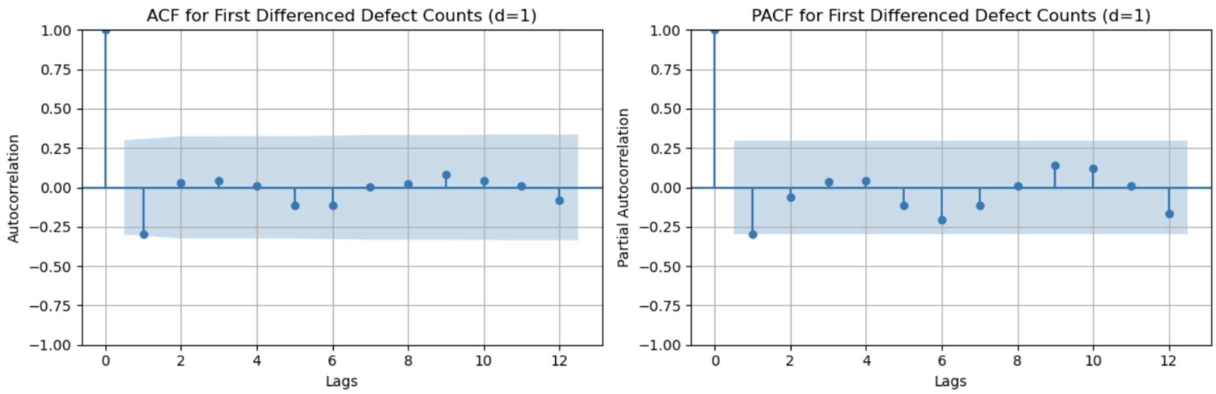
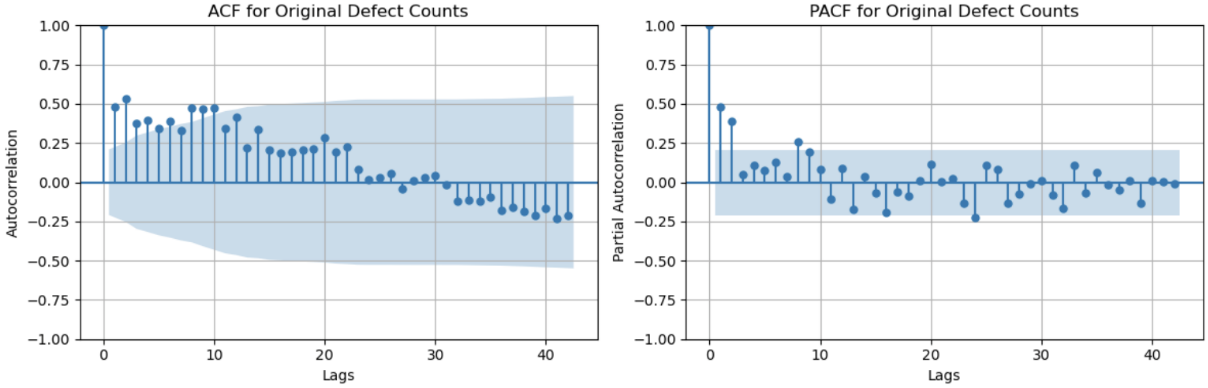


Figure D.8: ACF, PACF, and ADF for Part A1_Mh

```
ADF Statistic for Original Series: -1.412795134791986
p-value for Original Series: 0.5761268310993166
The time series Original Series is non-stationary (fail to reject the null hypothesis of the ADF
test).
```




```
ADF Statistic for First Differenced Series: -4.946333700430672
p-value for First Differenced Series: 2.8311617632037945e-05
The time series First Differenced Series is stationary (reject the null hypothesis of the ADF tes
t).
```

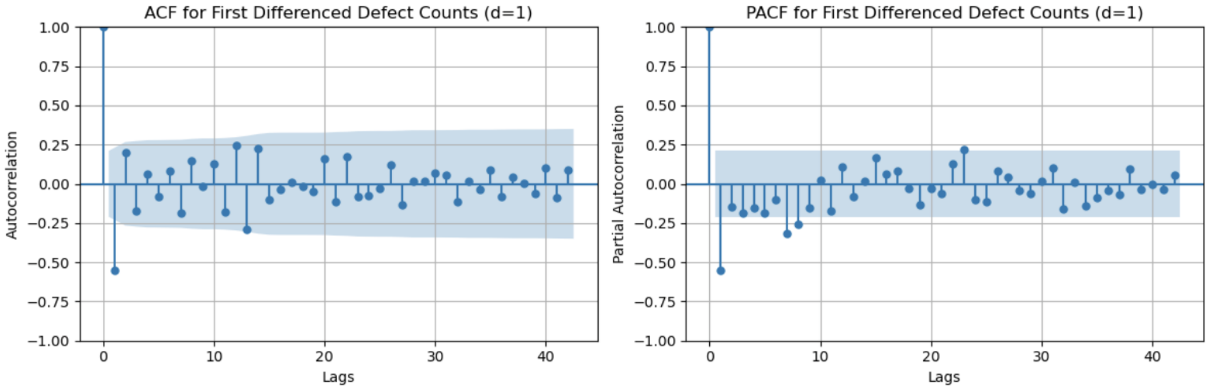
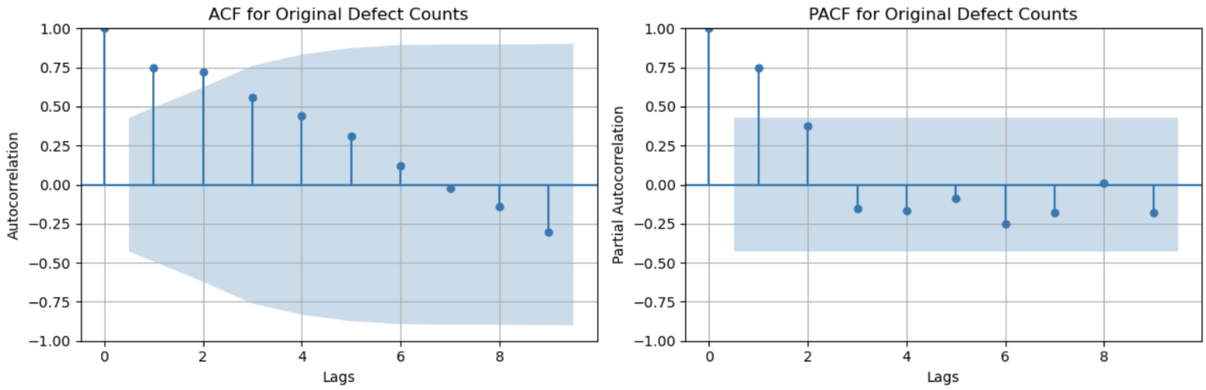


Figure D.9: ACF, PACF, and ADF for Part A2_Wk

```
ADF Statistic for Original Series: -1.2283183213159405
p-value for Original Series: 0.6612968228012319
The time series Original Series is non-stationary (fail to reject the null hypothesis of the ADF
test).
```




```
ADF Statistic for First Differenced Series: -8.480036865301477
p-value for First Differenced Series: 1.4011954647280645e-13
The time series First Differenced Series is stationary (reject the null hypothesis of the ADF tes
t).
```

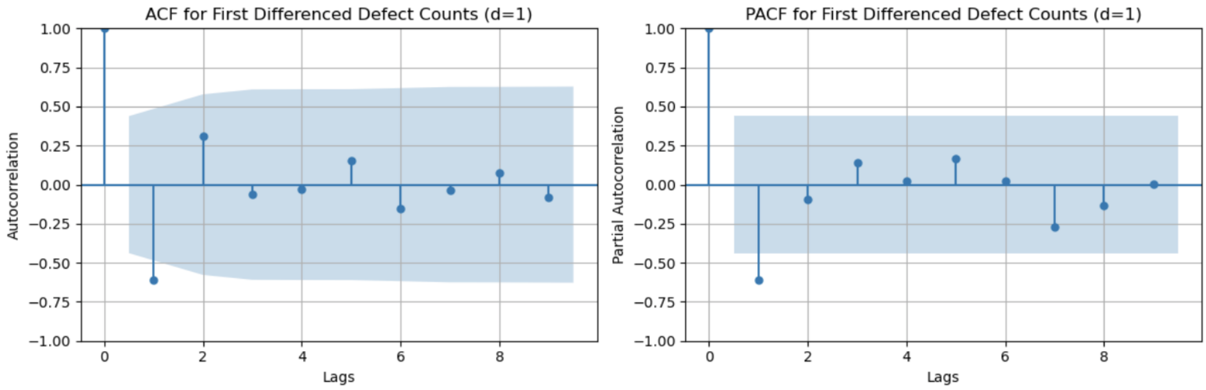
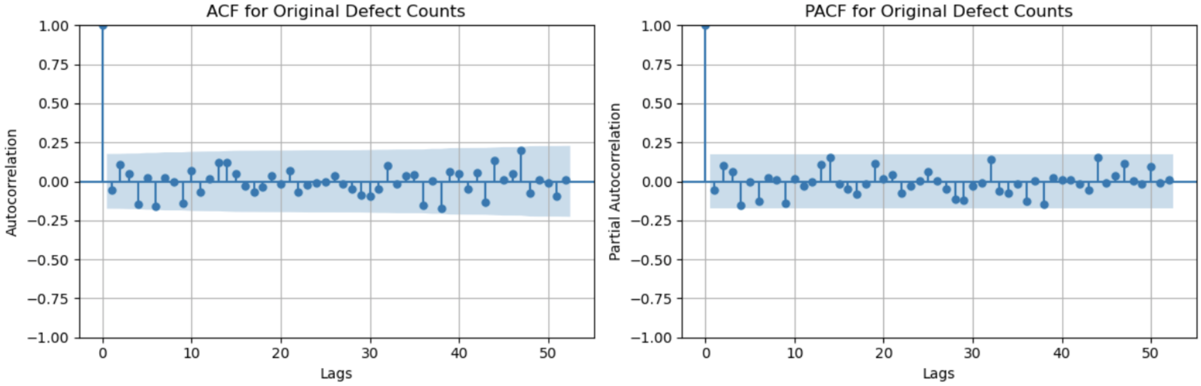


Figure D.10: ACF, PACF, and ADF for Part A2_Mh

```
ADF Statistic for Original Series: -5.7356914059206545
p-value for Original Series: 6.446412467084165e-07
The time series Original Series is stationary (reject the null hypothesis of the ADF test).
```




```
ADF Statistic for First Differenced Series: -6.517192657206735
p-value for First Differenced Series: 1.0640687871707964e-08
The time series First Differenced Series is stationary (reject the null hypothesis of the ADF tes
t).
```

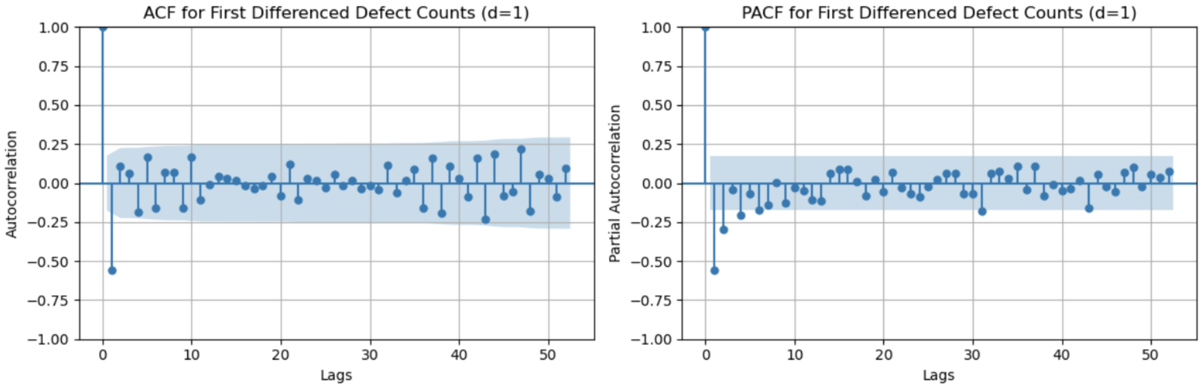
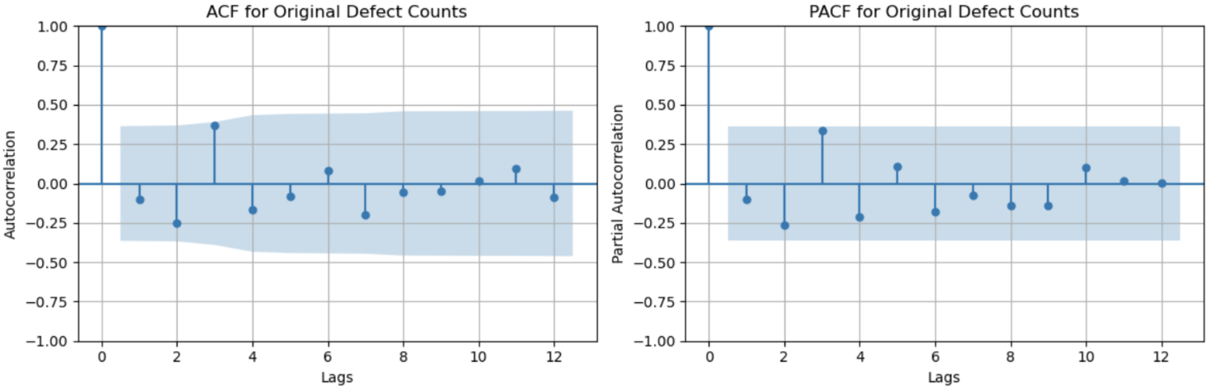


Figure D.11: ACF, PACF, and ADF for Part A3_Wk

```
ADF Statistic for Original Series: -2.2749361820424587
p-value for Original Series: 0.18017899778947716
The time series Original Series is non-stationary (fail to reject the null hypothesis of the ADF
test).
```




```
ADF Statistic for First Differenced Series: -9.388673147190026
p-value for First Differenced Series: 6.651766992942507e-16
The time series First Differenced Series is stationary (reject the null hypothesis of the ADF tes
t).
```

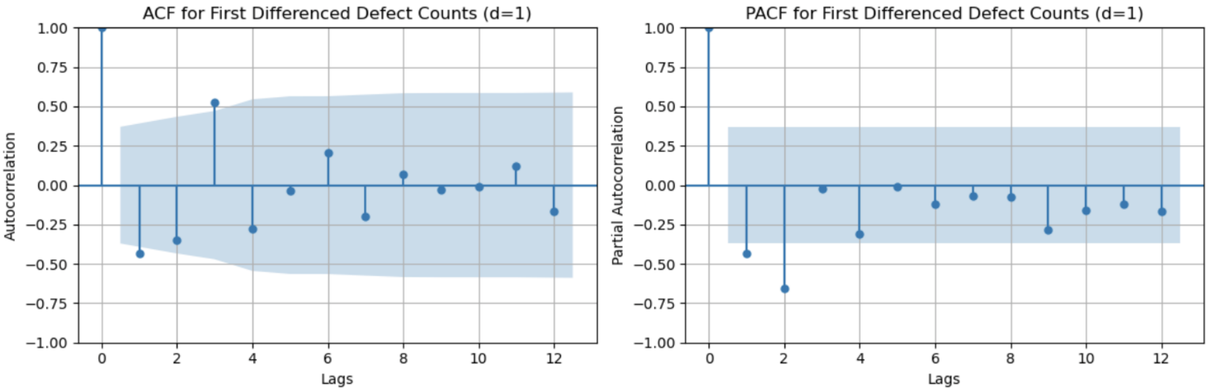
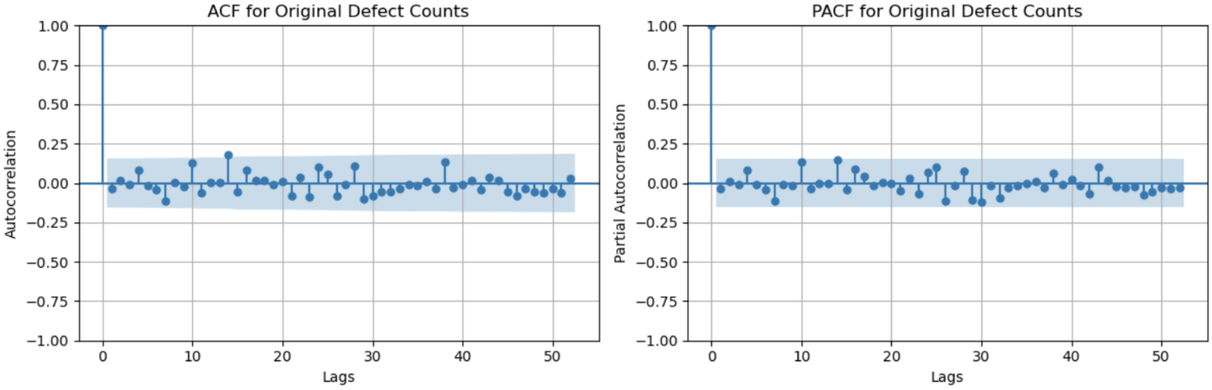


Figure D.12: ACF, PACF, and ADF for Part A3_Mh

```
ADF Statistic for Original Series: -14.123596896805086
p-value for Original Series: 2.400997544399026e-26
The time series Original Series is stationary (reject the null hypothesis of the ADF test).
```




```
ADF Statistic for First Differenced Series: -7.647799130299852
p-value for First Differenced Series: 1.8229312402200974e-11
The time series First Differenced Series is stationary (reject the null hypothesis of the ADF tes
t).
```

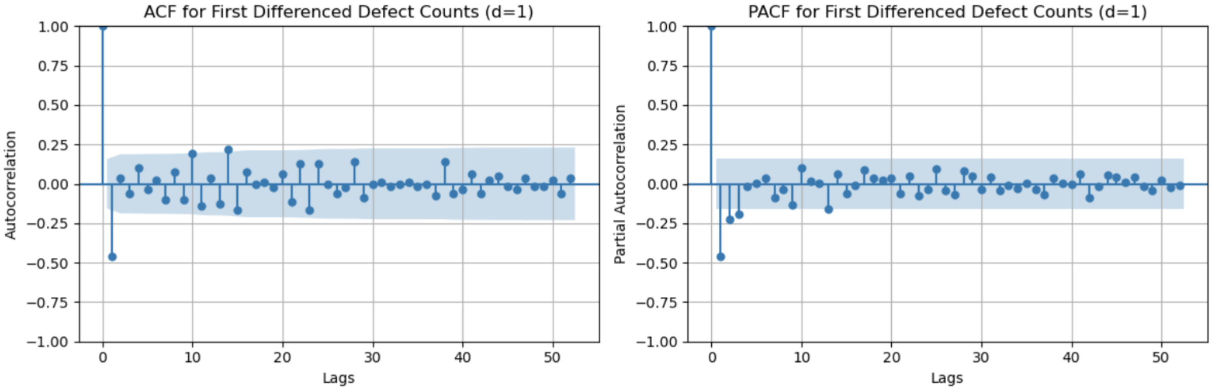
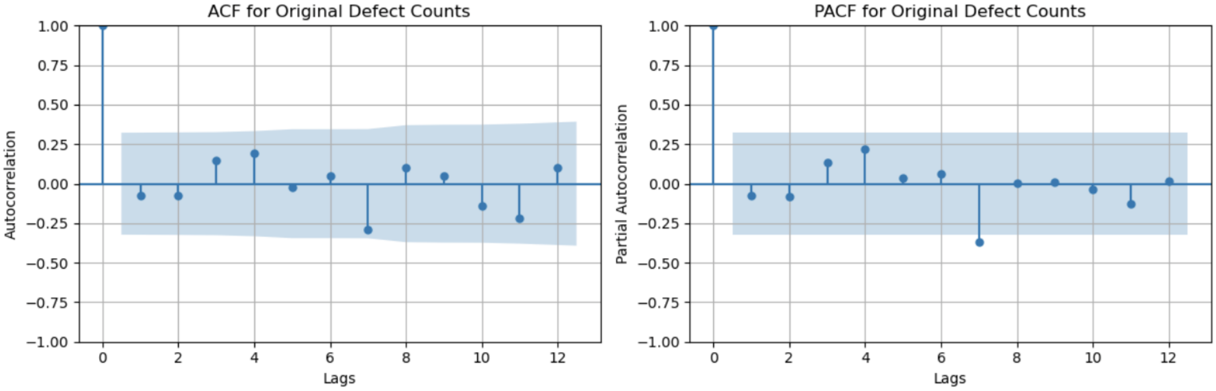


Figure D.13: ACF, PACF, and ADF for Part B1_Wk

```
ADF Statistic for Original Series: -6.641131663139212
p-value for Original Series: 5.4064416050185675e-09
The time series Original Series is stationary (reject the null hypothesis of the ADF test).
```




```
ADF Statistic for First Differenced Series: -6.765538459057994
p-value for First Differenced Series: 2.7231893158221562e-09
The time series First Differenced Series is stationary (reject the null hypothesis of the ADF tes
t).
```

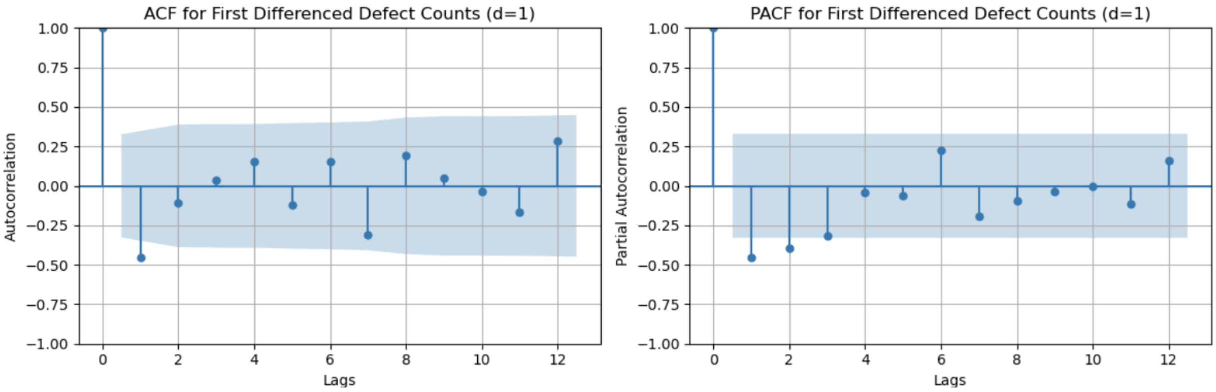
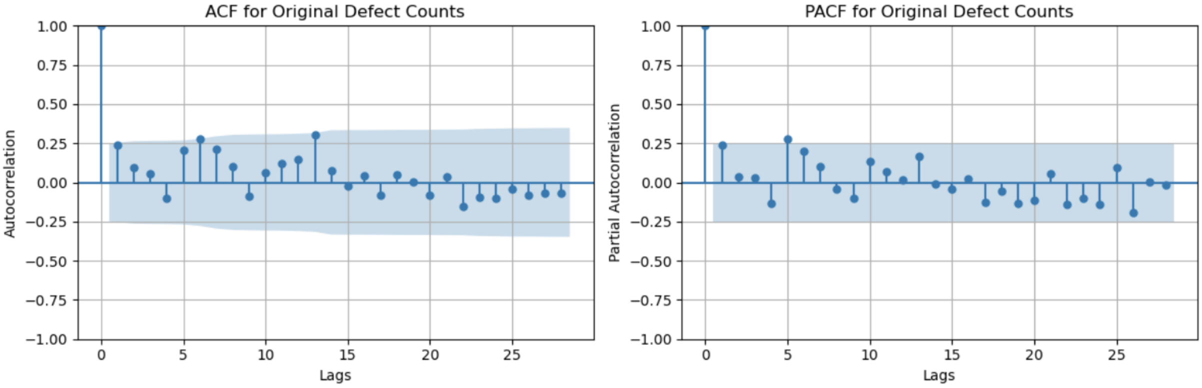


Figure D.14: ACF, PACF, and ADF for Part B1_Mh

```
ADF Statistic for Original Series: -5.9842247011314535
p-value for Original Series: 1.8081948263554375e-07
The time series Original Series is stationary (reject the null hypothesis of the ADF test).
```




```
ADF Statistic for First Differenced Series: -6.7908236356241485
p-value for First Differenced Series: 2.3671889135822982e-09
The time series First Differenced Series is stationary (reject the null hypothesis of the ADF tes
t).
```

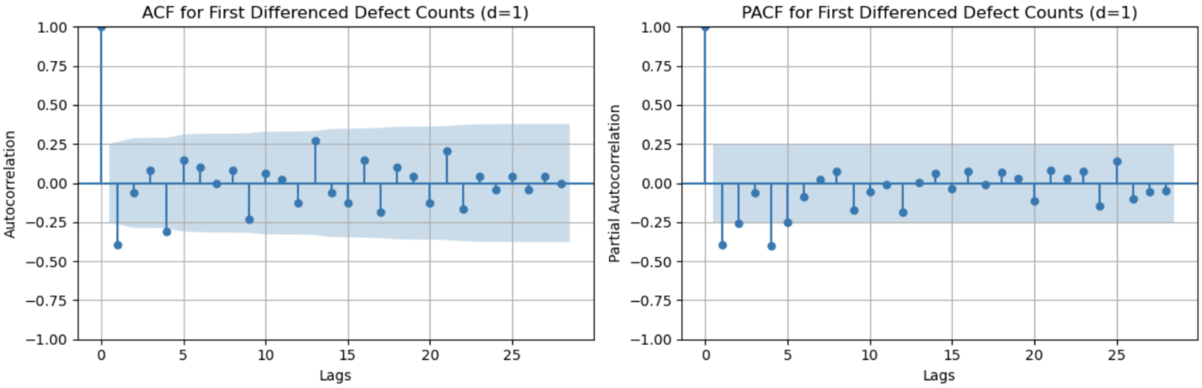
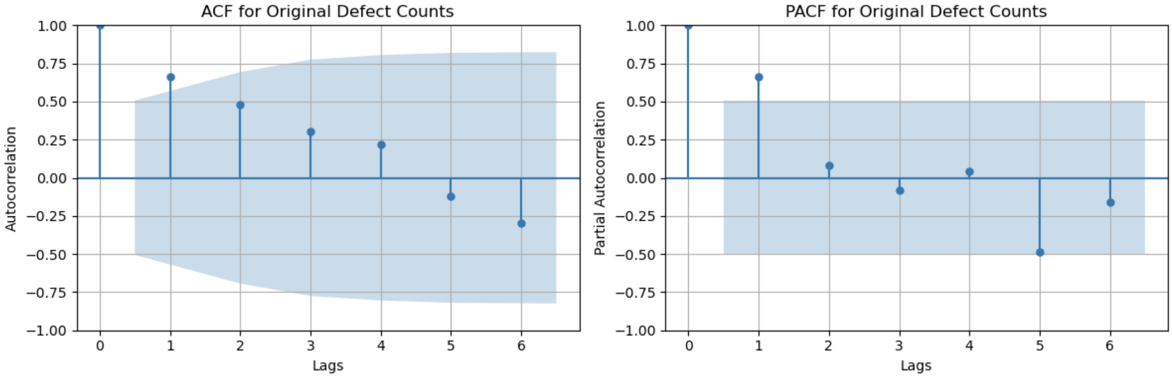
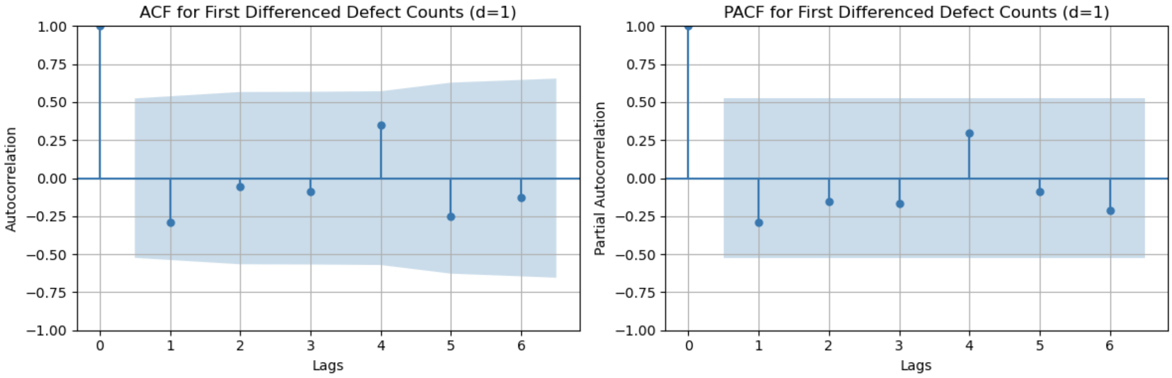


Figure D.15: ACF, PACF, and ADF for Part B2_Wk

```
ADF Statistic for Original Series: -2.1273957875114324
p-value for Original Series: 0.23364223705285242
The time series Original Series is non-stationary (fail to reject the null hypothesis of the ADF test).
```




```
ADF Statistic for First Differenced Series: -0.9042941590232915
p-value for First Differenced Series: 0.7865550002645437
The time series First Differenced Series is non-stationary (fail to reject the null hypothesis of the ADF test).
```




```
ADF Statistic for Second Differenced Series: -4.35772396978005
p-value for Second Differenced Series: 0.00035236450377913324
The time series Second Differenced Series is stationary (reject the null hypothesis of the ADF test).
```

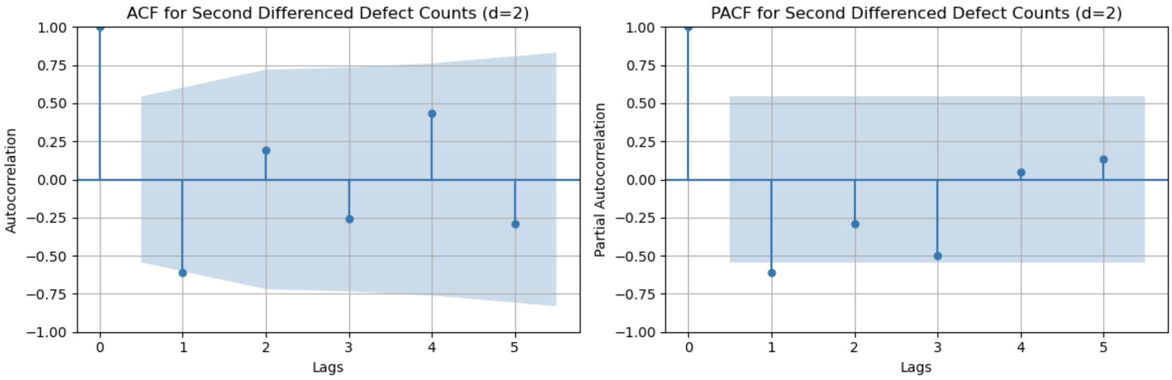
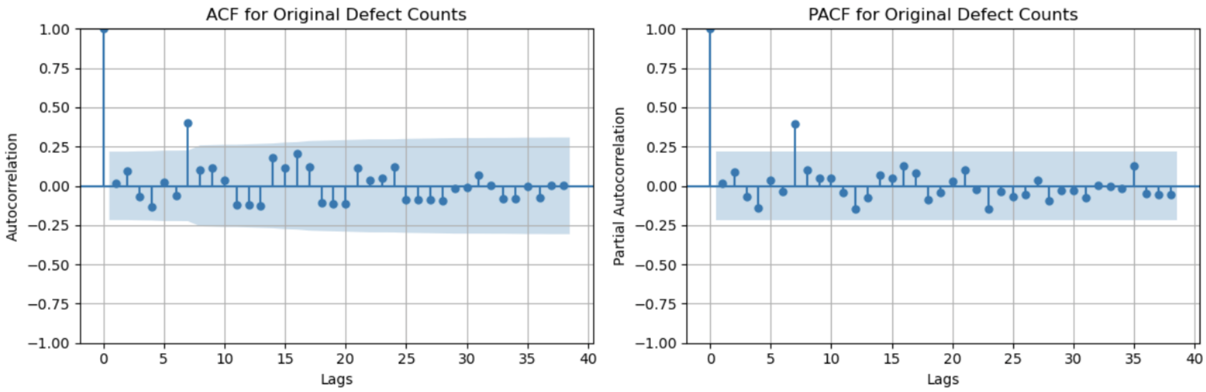


Figure D.16: ACF, PACF, and ADF for Part B2_Mh

```
ADF Statistic for Original Series: -2.1414798962945167
p-value for Original Series: 0.2281655929814851
The time series Original Series is non-stationary (fail to reject the null hypothesis of the ADF
test).
```




```
ADF Statistic for First Differenced Series: -7.13274937679553
p-value for First Differenced Series: 3.4834003780816843e-10
The time series First Differenced Series is stationary (reject the null hypothesis of the ADF tes
t).
```

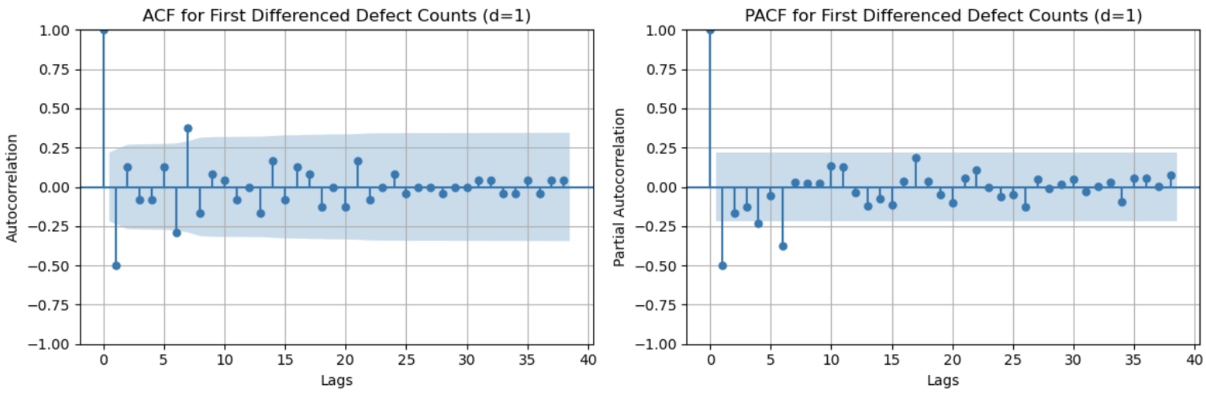
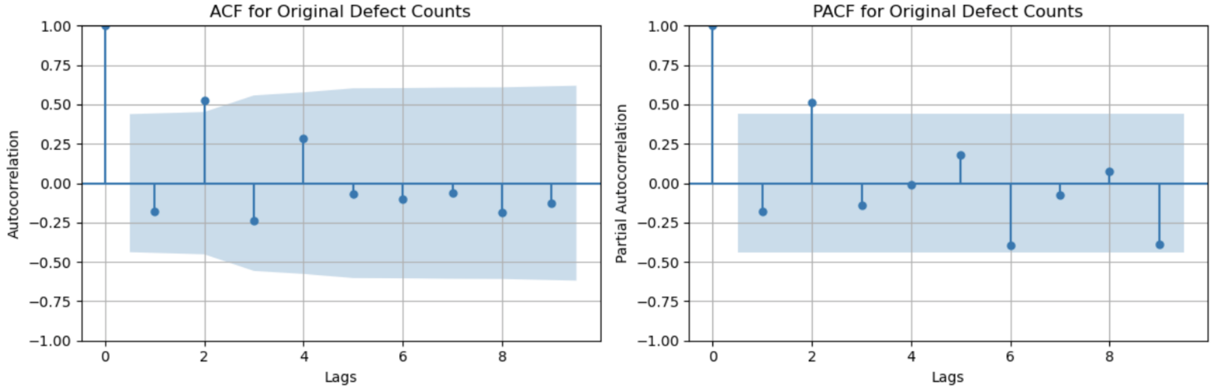


Figure D.17: ACF, PACF, and ADF for Part C_Wk

```
ADF Statistic for Original Series: -3.8487613337743083
p-value for Original Series: 0.002446023684053248
The time series Original Series is stationary (reject the null hypothesis of the ADF test).
```




```
ADF Statistic for First Differenced Series: -12.12436085996397
p-value for First Differenced Series: 1.794873871787302e-22
The time series First Differenced Series is stationary (reject the null hypothesis of the ADF tes
t).
```

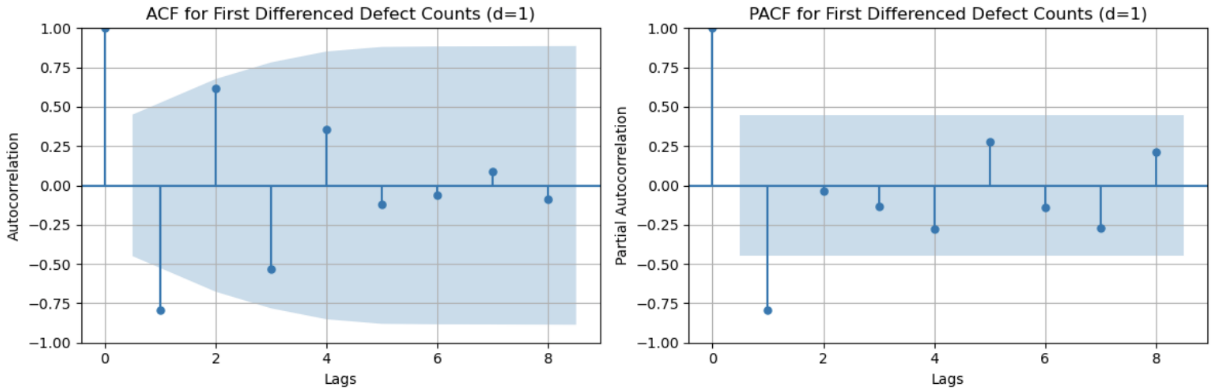


Figure D.18: ACF, PACF, and ADF for Part C_Mh

D.4. Weekly Forecasting Results Models Comparison

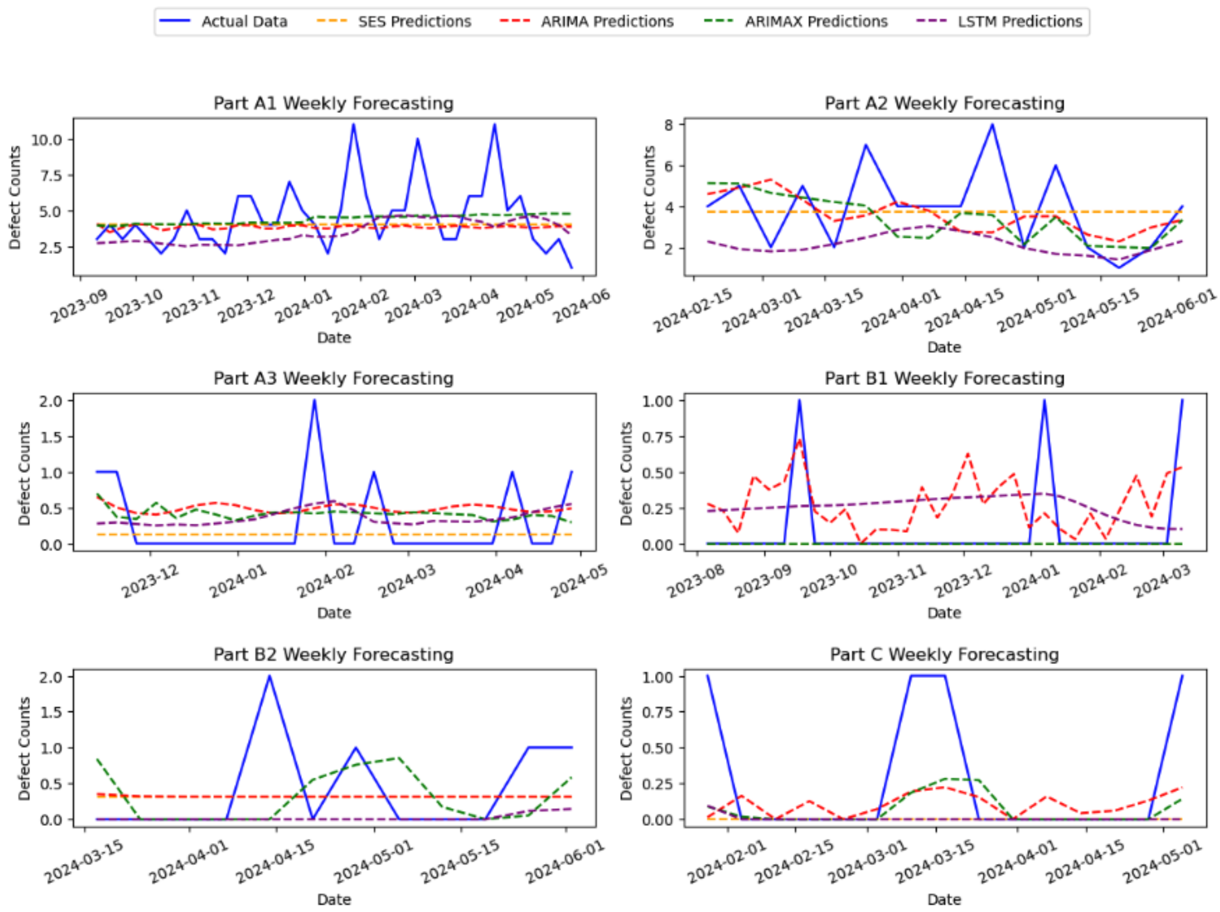


Figure D.19: Weekly SES, ARIMA, ARIMAX, LSTM Results

D.5. Monthly Forecasting Results Models Comparison

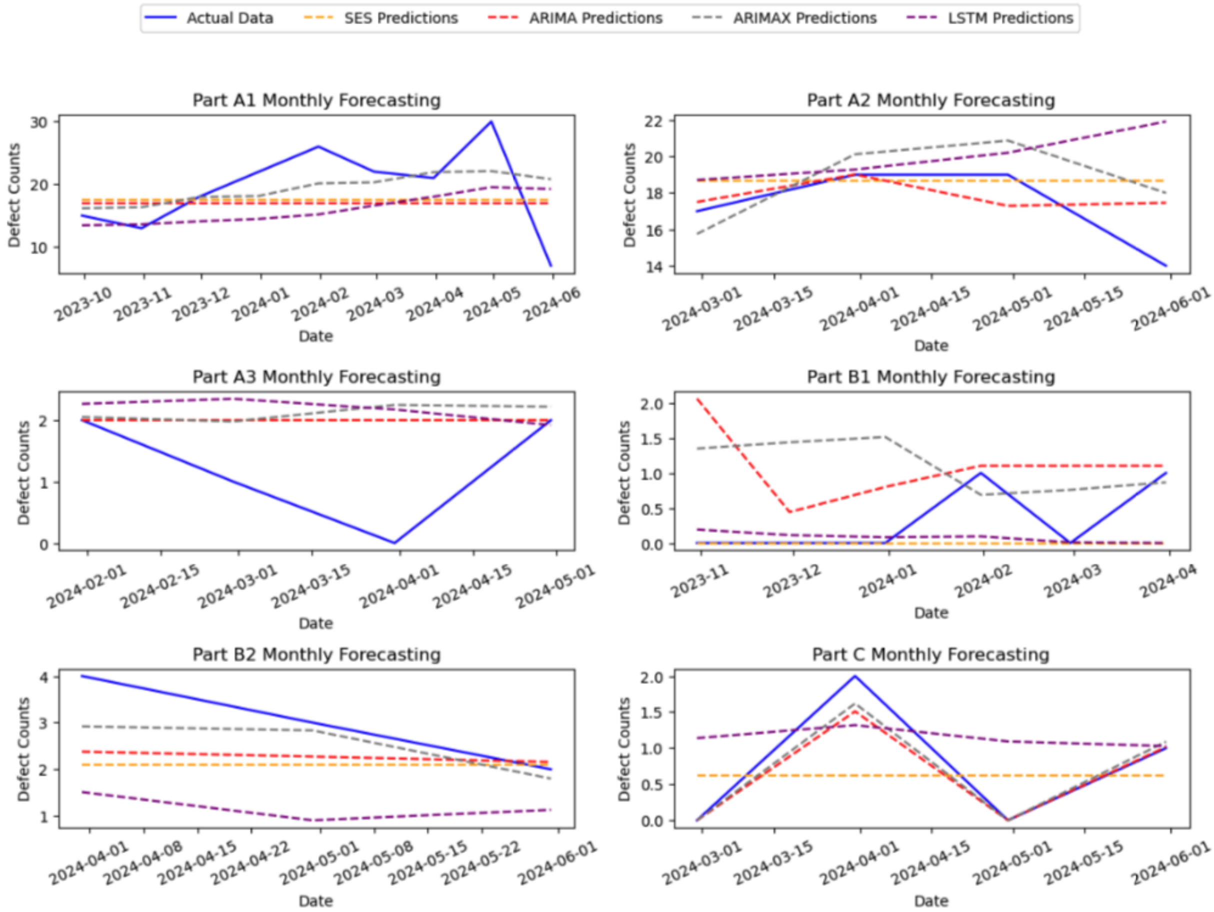


Figure D.20: Monthly SES, ARIMA, ARIMAX, LSTM Results